بسم الله الرحمن الرحيم

كتاب الاحتمال والإحصاء

نسخة إلكترونية مهداة من

مكتبة السماح للخدمات الجامعية ((أبو وائل))

لطلب نسخة منسوخة في حال عدم توفر النسخة الأصلية

يرجى الاتصال على 0933467204

أو تشريفنا على عنواننا

قريب من السكن الجامعي – مقابل حديقة أبناء الشهداء

أول جادة المعتصم 10م

لا تنسونا من صالح الدعاء

والله ولي التوفيق

Ayham Altaleb

الاحتمال والإحصاء

السنة: الثانية – الثالثة – الرابعة.

القسم: العلوم الأساسية.

الاختصاص: هندسة طاقة – هندسة الحواسيب والأتمتة – هندسة الغزل والنسيج.

منشورات جامعة دمشق
كلية الهندسة الميكانيكية والكهربائية

الاحتمال والإحصاء

الدكتور
علاء الدين القبانجي
أستاذ في قسم العلوم الأساسية

الدكتور
حسام حمامة كمرجي
مدرس في قسم العلوم الأساسية

جامعة دمشق

١٤٣٢-١٤٣٣ هـ
٢٠١١ - ٢٠١٢ م

فهرس

الموضوع	الصفحة
المقدمة	٩
الفصل الأول: مقدمة في الإحصاء الوصفي	**١١ـ٥٦**
1-1 – لمحة تاريخية	١٣
1-2 – تعريف الإحصاء والإحصاء الوصفي	١٥
1-3 – البيانات الإحصائية وطرق عرضها	١٥
1-4 – مقاييس النزعة المركزية والتشتت	٢٦
1-5 – المتغير المعياري	٤٤
1-6 – مبرهنة تشيبتشيف	٤٥
1-7 – القاعدة التجريبية	٤٨
1-8 – تمارين عامة	٥٣
الفصل الثاني: مبادئ الاحتمال	**٥٧ـ١٠٠**
2-1 – فضاء التجربة	٥٩
2-2 – الحدث	٦٢
2-3 – جبر الحوادث	٦٦
2-4 – الفضاء الاحتمالي والدالة الاحتمالية	٦٧
2-5 – خصائص الاحتمال	٧٢
2-6 – الاحتمال الشرطي	٧٩
2-7 – قانون بايز	٨٥
2-8 – طرائق العد	٨٩
2-9 – تمارين عامة	٩٥
الفصل الثالث: المتغيرات العشوائية	**١٠١ـ١٦٥**
3-1– مقدمة.	١٠٣
3-2 – المتغير العشوائي.	١٠٤

3-3 – التوزيع الاحتمالي.	١٠٦
3-4 – التوزيع المشترك لمتغيرين عشوائيين.	١١٧
3-5 – القيم المميزة للمتغيرات العشوائية.	١٢٧
3-6 – العزوم والدوال المولدة لها.	١٥٠
3-7 – تمارين عامة.	١٥٨
الفصل الرابع: التوزيعات الخاصة	**٢٠٧ـ١٦٧**
4-1 – التوزيع الثنائي	١٦٩
4-2 – توزيع بواسون	١٧٥
4-3 – التوزيع الطبيعي	١٧٩
4-4 – توزيع (ستيودنت) t.	١٨٩
4-2 – توزيع مربع كاي ،χ^2،.	١٩٥
4-6 – توزيع فيشر.	١٩٨
4-7 – تمارين عامة.	٢٠٦
الفصل الخامس: نظرية التقدير	**٢٥٢ـ٢٠٩**
5-1 – شروط التقدير.	٢١٢
5-2 – التقدير النقطي والمجالي للوسطاء.	٢١٥
5-3 – تقدير متوسط المجتمع الإحصائي μ.	٢١٨
5-4 – تقدير تباين المجتمع الإحصائي σ^2.	٢٣٣
5-5 – تقدير النسبة بين تبايني مجتمعين إحصائيين $\frac{\sigma_1^2}{\sigma_2^2}$.	٢٣٩
5-6 – التمثيل الهندسي (البياني) لمجال الثقة.	٢٤٢
5-7– تقدير الاحتمال P للمجتمع الإحصائي.	٢٤٣
5-8 – تقدير وسيط الفرق بين نسبتين P_D لمجتمعين إحصائيين.	٢٤٥
5-9 – تمارين عامة.	٢٤٩
الفصل السادس: اختبار الفرضيات	**٢٩٥ـ٢٥٣**
6-1– تمهيد.	٢٥٥
6-2 – اختبارات حول المتوسط (μ).	٢٦١
6-3– اختبارات حول التباين (σ^2).	٢٦٨

6-4 – اختبارات النسب والاحتمال.	٢٧٣
6-5 – اختبار التوافق والتجانس.	٢٨٠
6-6 – اختبار الاستقلالية والترابط.	٢٨٦
6-7 – اختبار الإشارة.	٢٨٩
6-8 – تمارين عامة.	٢٩٣
الفصل السابع: الانحدار والارتباط البسيط	**٣١٦ـ٢٩٧**
7-1 – الانحدار الخطي البسيط.	٣٠٢
7-2 – الارتباط الخطي البسيط ومعامل بيرسون.	٣٠٧
7-3 – ارتباط الرتب ومعامل سبيرمان.	٣٠٩
7-4 – معامل كندال لارتباط الرتب (τ).	٣١٢
7-5 – تمارين عامة.	٣١٥
الفصل الثامن: الأرقام القياسية والسلاسل الزمنية	**٣٥٩ـ٣١٧**
8-1 – مكونات السلاسل الزمنية.	٣٢٠
8-2 – نماذج السلاسل الزمنية.	٣٢٢
8-3 – التوجه العام (أو النمط) الخطي للسلسلة.	٣٢٣
8-4 – التوجه العام T (أو النمط) غير الخطي.	٣٢٩
8-5 – طريقة المتوسطات لتعديل السلسلة الزمنية.	٣٣٤
8-6 – حساب وتقدير مركبة المؤثرات الموسمية.	٣٣٧
8-7 – حساب وتقدير مركبة المؤثرات الدورية.	٣٤٣
8-8 – الأرقام القياسية.	٣٤٧
8-9 – تمارين عامة.	٣٥٧
الفصل التاسع: نظرية ضبط الجودة	**٤٢١ـ٣٦١**
9-1 – تمهيد.	٣٦٣
9-2 – خرائط المراقبة وضبط الجودة.	٣٦٤
9-3 – تكوين خرائط المراقبة وضبط الجودة.	٣٦٨
9-4 – تحليل خرائط الضبط.	٣٧٢
9-5 – خرائط ضبط المتوسط $\bar{X}$.	٣٧٧

9-6 – خرائط ضبط المدى R.	٣٨٢
9-7 – خرائط ضبط المدى والمتوسط المشتركة.	٣٨٥
9-8 – خرائط ضبط الانحراف المعياري والمتوسط المشتركة.	٣٨٦
9-9 – خرائط ضبط النسبة P.	٣٨٧
9-10 – خرائط أعداد العيوب C.	٣٩٩
9-11 – خرائط أعداد العيوب u للوحدات الإنتاجية.	٤٠٣
9-12 – توزيعات باريتو.	٤٠٧
9-13 – مخططات الأسباب وتأثيراتها.	٤١٠
9-14 – قابلية أو قدرة العملية الإنتاجية.	٤١٢
9-15 – تمارين عامة.	٤١٧

المقدمة

لقد وضع هذا الكتاب كمنهج تطبيقي مخصص لطلبة كليات الهندسة الميكانيكية والكهربائية الذين لا تتعدى خلفيتهم الدراسية منهاج الدراسة الثانوية. ونظراً لهذا التوجه العام فقد رجّحنا الانحياز للمسائل والمضامين الهندسية بدرجة عالية خصوصاً في موضوعات اختبار الفرضيات والسلاسل الزمنية ومراقبة الجودة الإنتاجية.

وعلى الرغم من شعورنا بأنّ موضوع الإحصاء التطبيقي لا يمكن حصره في كتيّب منهجي واحد إلاّ أننا على قناعة تامة بأنّ ما استطعنا حصره بين دفتي هذا المنهاج المقرّر يصلح ويغطي الزمن المشروط المخصص لتدريس هذه المادة للطلبة المعنيين. فلقد ضمّ المنهاج في فصوله الثلاثة الأولى كلاً من الإحصاء الوصفي والاحتمالات والمتغيرات العشوائية. أمّا الفصل الرابع فنجد فيه استعراضاً لبعض التوزيعات الاحتمالية الهامة يليه التحليلات الإحصائية للتقدير النقطي والمجالي واختبار الفرضيات في الفصل الخامس والسادس على التتالي. بينما تم تخصيص الفصول السابع والثامن والتاسع للموضوعات ذات المضامين التطبيقية التخصصية كالسلاسل الزمنية وخرائط المراقبة وضبط الجودة، كما تضمّن الكتاب تمارين محلولة ذات طابع تطبيقي تخصصي مقصود.

ولمّا كانت هذه هي الطبعة الأولى للمنهاج وعلى الرغم من مراعاتنا لمعادلة أكثر المعلومات في أقل الصفحات المحددة وأقصر الزمن، إلاّ إننا نأمل من مدرسي هذه المادة موافاتنا بمقترحاتهم وأمثلتهم التطبيقية ليتسنى لنا إدراجها في الطبعات اللاحقة.

المؤلفان

الفصل الأول

مقدمة في الإحصاء الوصفي

1-1 – لمحة تاريخية

1-2 – تعريف الإحصاء والإحصاء الوصفي

1-3 – البيانات الإحصائية وطرق عرضها

1-4 – مقاييس النزعة المركزية والتشتت

1-5 – المتغير المعياري

1-6 – مبرهنة تشيبتشيف

1-7 – القاعدة التجريبية

1-8 – تمارين عامة

مقدمة في الإحصاء الوصفي
Introduction to Descriptional Statistics

1-1- لمحة تاريخية:

لقد وردت كلمة إحصاء في اللغة العربية في الكثير من الآيات الكريمة كقوله تعالى ﴿وإن تعدّوا نعمةَ الله لا تحصوها﴾ (سورة ابراهيم، الآية:٣٤ - وسورة النمل، الآية:١٨) وقوله ﴿وكل شيء أحصيناه في إمامٍ مبين﴾ (سورة يس، الآية:١٢) وقوله ﴿وأحصى كلَّ شيءٍ عدداً﴾ (سورة الجن، الآية:٢٨) وكذلك قوله ﴿ولقَد أحصاهُم وعدَّهم عدّاً﴾ (سورة مريم، الآية:٩٤).
وأحصى هنا بمعنى عدَّ وحصر.

وتعني كلمة إحصاء باللغة اللاتينية «status» الحالة أو الوضع والتي شُكل منها فيما بعد عدة كلمات وبخاصة كلمة «stato» التي تعني الدولة للإشارة إلى المعلومات عن تركيب القضايا العامة وحالتها. وفي الفرنسية يعود أصل كلمة إحصاء «statistique» إلى كلمة «Etat» وتعريبها دولة حيث يوجد لها تفسيران الأول يتعلق بجمع المعلومات العددية عن الشؤون التي تهم الدولة وأملاكها من أجل تشكيل الجيوش والضرائب وإدارة الأملاك، والتفسير الثاني هو إجراء حساب «compte rendu» على شكل جداول.

لقد ظهرت كلمة إحصاء في ألمانيا في القرن السابع عشر، وتعني بالألمانية «statistik» وهي مشتقة من كلمة «staat» التي تعني الدولة، وكذلك الأمر بالنسبة للإنكليزية فكلمة إحصاء بالإنكليزية هي «statistics» والتي تعني إشراف الدولة وقد اشتقت من الكلمة «State» التي تعني الدولة. ومن هنا فالإحصاء هو مجموعة المعلومات عن الدولة. لقد ظهر الإحصاء وتطور بفعل حاجات المجتمع. واستخدم الإحصاء في البدء فقط لتعداد السكان وحصر مواردهم وغنائمهم وجنودهم وسلاحهم ومعرفة الثروات، إلخ...

لقد قام القدماء بأعمال التعداد الإحصائي للحصول على هذه المعلومات المختلفة فبالنسبة إلى تعداد السكان هناك معلومات عن تعداد السكان في مصر القديمة يعود للسنة 3500 ق.م وفي غيرها من مختلف دول عصر الرق. فقد كان لدى الصينيين، فيما بين الألفين الرابع والثاني ق.م معلومات عن السكان وكانوا يستعملون جداول إحصائية تتعلق بالزراعة. والشيء نفسه بالنسبة للمصريين الذين عرفوا التعداد الدائم ووضعوا أقدم «ميزان اقتصادي» عُرف في أيامهم وهو «ميزان النيل» فمستوى ارتفاع فيضان النيل كان مؤشراً ممتازاً للخصب، ويستعمل لتقدير حجم الضرائب. وكان الخليفة المأمون أكثر الخلفاء عنايةً واهتماماً باستخدام طريقة البيانات العددية والاستعانة بها في تنظيم أمور الدولة.

وفي روما القديمة كان الغرض من التعداد هو تعيين حجم الضريبة المتوجب دفعها للدولة الرومانية. لقد بقيت الطريقة الإحصائية وصفية حتى القرن الثامن عشر حيث بدأت تبرز مدرستان متميزتان: المدرسة الوصفية الألمانية ومدرسة الحسابين السياسيين. فيما بعد وبالضبط بعد ظهور نظرية الاحتمالات ارتبط علم الإحصاء بأسماء كل من برنولي (Jakob Bernoulli, 1654-1705) مكتشف قانون الأعداد الكبيرة ودي موافر (A,deMoiver,1667-1754) ولابلاس (P.Laplace, 1749-1827) وغاوص (C.F.Gauss, 1777-1855) وبواسون (S.D. Poisson, 1781-1840) وكتيليه (Quetelet, 1796-1894) الذي يعود إليه الفضل في نقل الأساليب الإحصائية الجديدية إلى الاقتصاد والديموغرافية.

وفي نهاية القرن التاسع عشر وبداية القرن العشرين وبعد تطور علم الاحتمالات على أيدي كل من تشيبتشيف (P.L.Tschebyscheff, 1821-1894) وماركوف (A.A.Markov, 1856-1922) وكولموغوروف (A.N.Kalmogorov) وغينيدنكو (B.W.Gnedenko) تطور علم الإحصاء وأخذ يتقدم بخطوات سريعة وأصبح يطبق في شتى العلوم وأصبحت الدول المتقدمة تعد الإحصاء وسيلة هامة في إدارة القرارات واتخاذها في الأمور الصحية والعسكرية والمالية والإدارية، وقد ساهم كبار العلماء في إيجاد الطرائق الكفيلة بتطوير هذا العلم واستخداماته في مختلف المجالات أمثال بيرسون (K.Pearson, 1857-1936) الذي اتخذ من علم الوراثة ميداناً لتجاربه ولاختبار نظرياته بالإضافة إلى مندل المشهور بنظرياته في البيولوجيا كما قام فيشر (R.A.Fisher, 1890-1962) وولد (A.Wald, 1902-1950) ونويمان (J. Neyman)

وستودنت (Student) وكندال (Kendall) وسميرنوف (N.W.Smirnow) ورومانوفسكي (W.J.Romanowski) بتقديم جهود ساهمت بتقدم علم الإحصاء.

1-2- تعريف الإحصاء والإحصاء الوصفي:

الإحصاء: هو علم يهتم بتجميع وتنظيم وتحليل وتفسير ظاهرة معينة لمجتمع إحصائي ما، واستخلاص نتائج مقبولة لاعتمادها أساساً لفكرة واضحة ومعقولة عن تلك الظاهرة والسعي للوصول إلى قرار مبني على البيانات الإحصائية التي قام بدراستها الإحصائي. وبالتالي فإن للإحصاء دوراً أساسياً في التطوير الصناعي والعلمي والاجتماعي حيث يستخدم في جميع العلوم الطبية والاجتماعية والاقتصادية والهندسية والفيزيائية والكيميائية وغيرها..

إن هذا التعريف يقودنا إلى التمييز بين وظيفتين للطريقة الإحصائية: التقنيات الوصفية الإحصائية، وتقنيات الاستقراء الإحصائي. إن الاهتمام الأكبر للإحصاء الوصفي ينصب على تهيئة المعلومات وإعدادها في شكل مناسب قابل للاستخدام والفهم. أمّا الاستقراء الإحصائي، فيهتم بتعميم هذه المعلومات، وبصورة خاصة، يهتم بالقيام باستقراءات حول المجتمع الأصلي تستند إلى عينات مأخوذة منه.

وبالتالي يمكننا تعريف **الإحصاء الوصفي** على أنه عملية جمع المعطيات الإحصائية أو المعلومات ومن ثم تبويب هذه المعطيات في فئات متجانسة وفي جداول ورسوم بيانية، بالإضافة إلى حساب بعض الخصائص العددية الهامة لهذه المعطيات.

1-3- البيانات الإحصائية وطرق عرضها:

عند دراسة ظاهرة معينة نلاحظ أن المعطيات الإحصائية المجمعة غالباً ما تكون غير مرتبة. لذلك هناك طرق معينة تقوم بترتيب هذه البيانات وعرضها بشكلٍ جديد يمكننا من دراسة تلك الظاهرة بشكلٍ أفضل والوصول للنتائج المطلوبة من خلال استقراء تلك البيانات وتحليلها.

1-3-1- جدول التوزيع التكراري (Frequency distribution table):

ليكن لدينا مجموعة من القياسات (المشاهدات – البيانات) لظاهرة معينة، عندئذٍ يمكننا ترتيب هذه القياسات وعرضها ضمن جدول حيث تكفي نظرة سريعة إلى هذا الجدول لإعطاء فكرة واضحة عن الظاهرة المدروسة.

وهناك نوعان من الجداول، **جدول التوزيع التكراري النقطي** وهو جدول يتم فيه كتابة القياسات بشكل مرتب (تصاعدياً أو تنازلياً) مع كتابة التكرار لكل قياس على حدى (عدد مرّات ظهور هذا القياس). وكذلك هناك **جدول التوزيع التكراري للبيانات المبوبة*** وهو جدول يتم فيه تصنيف القياسات وتبويبها ضمن فئات مع ذكر التكرار لكل فئة (عدد القياسات الموجودة ضمن الفئة الواحدة). ويُحدد طول الفئة من قبل الإحصائي حيث يجب أن يكون عدد هذه الفئات معقولاً (عادةً يُؤخذ عدد الفئات بين 5 و 20 فئة). وسنقتصر في دراستنا على الحالة التي تكون فيها الفئات متساوية الطول.

يُستخدم عادةً جدول التوزيع التكراري النقطي عندما يكون عدد القياسات غير كبير وفي الحالة المعاكسة أي عندما يكون عدد القياسات كبيراً يُستَخدم جدول التوزيع التكراري للبيانات المبوبة.

مثال (1) : البيان التالي يدل على كمية الكهرباء المستهلكة يومياً (بآلاف الكيلوواط) في إحدى المدن خلال فترة زمنية مقدارها 50 يوم متتالي:

25	49	43	28	30	38	34	48	39	37
38	28	34	38	48	25	39	37	43	38
39	43	38	43	28	43	37	39	48	30
43	30	39	34	37	39	49	34	43	38
43	25	43	37	39	38	48	43	38	34

والمطلوب:

1- أنشئ جدول التوزيع التكراري النقطي.

2- أنشئ جدول التوزيع التكراري للبيانات المبوبة.

الحل: 1- بملاحظة القياسات المبينة أعلاه نجد أن جدول التوزيع التكراري النقطي لهذه القياسات هو:

الجدول (1.1)- جدول التوزيع التكراري النقطي

القياسات	25	28	30	34	37	38	39	43	48	49
التكرار	3	3	3	5	5	9	6	10	4	2

2- نلاحظ من خلال هذه القياسات أن أكبر قيمة هي 49 وأصغر قيمة هي 25 فإذا اخترنا عدد الفئات مساوياً 5 أي طول كل فئة 5 عندئذٍ يكون جدول التوزيع للبيانات المبوبة هو الجدول التالي:

* يصطلح على تسمية جدول التوزيع التكراري للبيانات المبوبة اسم جدول التوزيع التكراري الفئوي.

الجدول (1.2)- جدول التوزيع التكراري للبيانات المبوبة

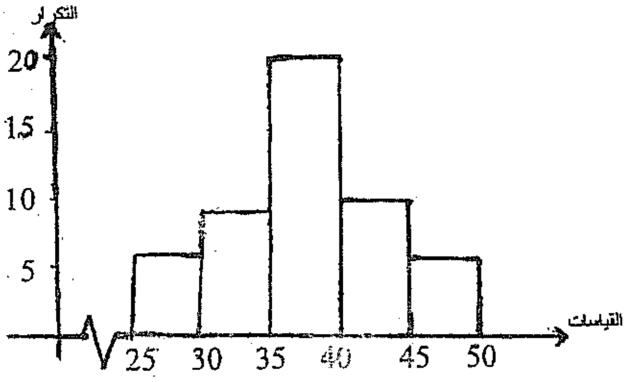
رقم الفئة	حدود الفئة	مركز الفئة	التكرار	التكرار النسبي	التكرار المئوي
1	[25,30[	27.5	6	0.12	12
2	[30,35[	32.5	8	0.16	16
3	[35,40[	37.5	20	0.40	40
4	[40,45[	42.5	10	0.20	20
5	[45,50[	47.5	6	0.12	12
المجموع			50	1	100

حيث نلاحظ أننا مثلنا كل فئة بفترة نصف مفتوحة كما ذكرنا مركز الفئة (لأهميته مستقبلاً) وهو يساوي نصف مجموع حدّي الفئة. وفي العمود الرابع سُجل التكرار لكل فئة. أمّا في العمودين الخامس والسادس فقد ذُكر التكرار النسبي والتكرار المئوي لكل فئة حيث نحصل على التكرار النسبي للفئة بتقسيم تكرار هذه الفئة على العدد الكلي للقياسات، كما نحصل على التكرار المئوي للفئة بضرب تكراراها النسبي بمئة.

ملاحظة ①: لاحظ أننا أخذنا أطوال الفئات متساوية وطول كل منها ℓ يساوي $\ell=\frac{U-L}{N}$ أي حاصل قسمة الفرق بين الحدين الأعلى والأدنى للبيانات على عدد الفئات N. حيث ندعو الفئات في هذه الحالة فئات منتظمة وفي الحالة التي لا يتساوى فيها أطوال الفئات تسمى الفئات بالفئات غير المنتظمة، وسوف ندرس حالة الفئات المنتظمة فقط.

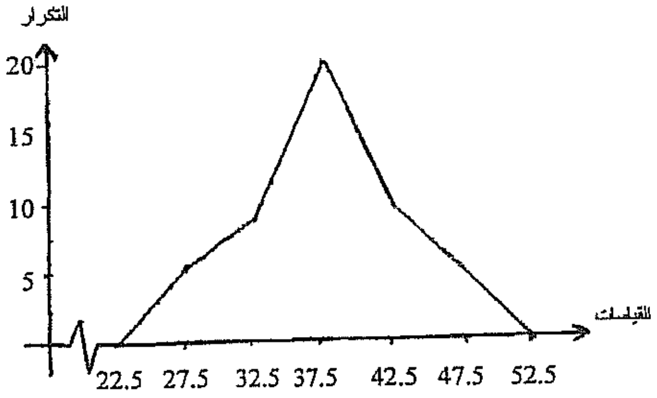
2-3-1- المضلع التكراري:

هناك طرق أخرى لعرض البيانات الإحصائية وذلك لسهولة التعامل معها، ومن هذه الطرق عرض البيانات على شكل مدرج تكراري «Histogram» حيث نمثل كل فئة بمستطيل قاعدته تساوي طول الفئة وارتفاعه يساوي تكرار تلك الفئة كما في الشكل التالي والذي يمثل المدرج التكراري للبيانات المبينة في الجدول (1.2):



الشكل (1.1): المدرج التكراري

كما يمكن عرض البيانات على شكل مضلع يسمى المضلع التكراري وهو خط منكسر يصل بين النقاط التي فواصلها مراكز الفئات وتراتيبها تكرار تلك الفئات كما في المضلع التكراري التالي والذي يمثل البيانات المبينة في الجدول (1.2).

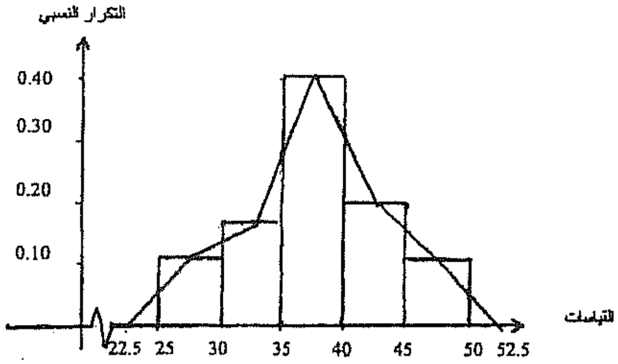


الشكل (1.2): المضلع التكراري

حيث أضفنا إلى الرسم القيمتين 22.5 و 52.5 في البداية والنهاية واللتين تمثلان المراكز للفئات التي تكرارها معدوم. لاحظ أنه يمكن رسم المضلع التكراري مباشرةً من المدرج التكراري وذلك بالوصل بين منتصفات القواعد العليا للمستطيلات السابقة بقطع مستقيمة. وقد يكون من الأنسب على الغالب تعديل المدرج التكراري بحيث يصبح ارتفاع المستطيل المنشأ فوق كل فئة مساوياً للتكرار النسبي الموافق لها وهو ما يسمى بالمدرج التكراري النسبي كما

أن المضلع الذي يصل بين منتصفات القواعد العليا لهذه المستطيلات يدعى بالمضلع التكراري النسبي.

فيما يلي المدرج (والمضلع) التكراري النسبي للبيانات المذكورة في الجدول (1.2):



الشكل (3-1): المدرج (والمضلع) التكراري النسبي

مما سبق نجد أن تمثيل جدول التوزيع التكراري للبيانات المبوبة بمدرج تكراري (أو تكراري نسبي) وكذلك بمضلع تكراري (أو تكراري نسبي) يمكننا من وصف البيانات بشكل أسرع وأدق. ففي مثالنا السابق نلاحظ من خلال الأشكال (1-1) أو (2-1) أو (3-1) أن استهلاك الكهرباء بمقدار أقل من 30 ألف كيلوواط أو بمقدار أكثر من 54 ألف كيلوواط هو شيء قليل نسبياً. بينما نرى أنه غالباً ما يكون الاستهلاك بين 30 ألف و 45 ألف كيلوواط يومياً.

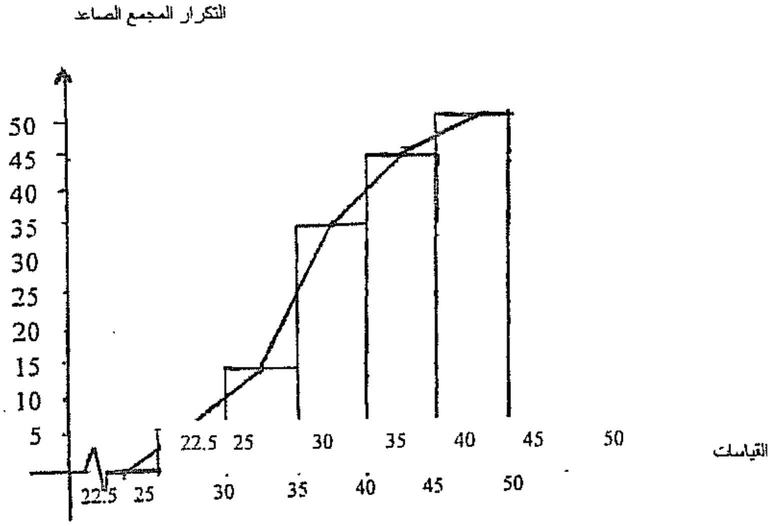
1-3-3- التوزيع التكراري المجمع:

هناك نوعان آخران لجداول التوزيع التكرارية وهما جدول التوزيع التكراري الصاعد وجدول التوزيع التكراري النازل والتي تدعى أيضاً جداول التوزيعات التكرارية المجمعة أو المتراكمة. فيما يلي جدول التوزيع التكراري المجمع الصاعد للبيانات المبينة في مثالنا السابق:

الجدول (3-1)- جدول التوزيع التكراري المجمع الصاعد

التكرار المجمّع الصاعد النسبي	التكرار المجمّع الصاعد	الحدود الدنيا للفئات
0	0	أقل من 25
0.12	6	أقل من 30
0.28	14	أقل من 35
0.68	34	أقل من 40
0.88	44	أقل من 45
1	50	أقل من 50

حيث يمثل التكرار المجمع الصاعد لكل فئة، مجموع تكرارات القيم الأصغر من الحد الأدنى لهذه الفئة. وبالتالي فإن المدرج (والمضلع) التكراري المجمع الصاعد الممثل لهذا الجدول يكون:

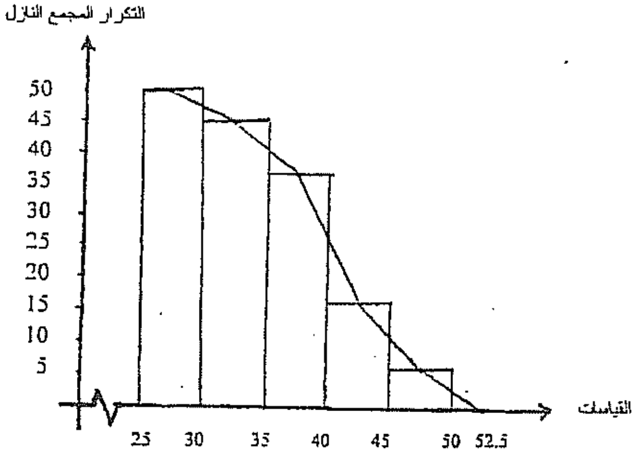


الشكل (4-1): المدرج (والمضلع) التكراري المجمع الصاعد

كذلك نبيّن فيما يلي جدول التوزيع التكراري المجمع النازل والمدرج (وبالتالي المضلع) التكراري المجمع النازل للبيانات المذكورة في مثالنا السابق حيث نبتدئ بالمجموع الكلي للبيانات.

الجدول (4-1)- جدول التوزيع التكراري المجمع النازل

الحدود العليا للفئات	التكرار المجمّع النازل	التكرار المجمّع النازل النسبي
أكبر أو يساوي 25	50	1
أكبر أو يساوي 30	44	0.88
أكبر أو يساوي 35	36	0.72
أكبر أو يساوي 40	16	0.32
أكبر أو يساوي 45	6	0.12
أكبر أو يساوي 50	0	0



الشكل (1-5): المدرج والمضلع التكراري المجمع النازل

حيث يمثل التكرار المجمع النازل لكل فئة، مجموع تكرارات القيم الأكبر من أو تساوي الحد الأعلى لهذه الفئة.

4-3-1- المجتمع الإحصائي والعينة (Population and Sample):

لو عدنا إلى مثالنا السابق نلاحظ أننا قمنا بتسجيل كمية الكهرباء المستهلكة خلال 50 يوم فقط وتعتبر هذه القياسات المسجلة عبارة عن عينة من مجتمع أيام السنة بكاملها. وبالتالي يمكننا تعريف المجتمع الإحصائي بأنه مجموعة كل العناصر التي تشملها دراسة إحصائية ما، بينما العينة هي مجموعة جزئية من المجتمع الإحصائي. فإذا كنّا نبحث في مسألة تحديد نسبة

المدخنين بين المهندسين في القطر العربي السوري فإن المجتمع الإحصائي لهذه المسألة الإحصائية هو جميع المهندسين في القطر السوري، بينما العينة في هذا المثال، فهي أي مجموعة من المهندسين السوريين. أمّا إذا كنا نبحث في مسألة تحديد نسبة المهندسين السوريين المدخنين الذين يفضلون نوعاً معيناً من الدخان فإن المجتمع الإحصائي هنا هو مجتمع المهندسين المدخنين فقط في سورية. وأي مجموعة مهندسين مدخنين جزئية من هذا المجتمع تمثل عينة. كذلك نلاحظ عند البحث في مسألة متوسط عمر محركات كهربائية منتجة في شركة صناعية معينة، يكون المجتمع الإحصائي في هذه الحالة هو كل المحركات المنتجة في هذه الشركة.

عند دراسة مسألة إحصائية على مجتمع ما غالباً ما تكون الدراسة صعبة وأحياناً مستحيلة، لذلك يلجأ الإحصائي إلى اختيار عينة من المجتمع المدروس حيث يقوم بتسجيل المشاهدات ودراستها للوصول إلى قرار يخص المجتمع بكامله. فمثلاً في مسألة تحديد نسبة المهندسين المدخنين في القطر السوري لا يمكننا القيام بسؤال كل مهندس سوري إذا كان يدخن أم لا، لذلك نكتفي عادةً بتوجيه السؤال لمجموعة من المهندسين تمثل عينة من مجتمع المهندسين السوريين وبناءً على النتائج التي نستخلصها من دراستنا لهذه العينة نستطيع إعطاء قرار حول نسبة المهندسين المدخنين يخص مجتمع المهندسين السوريين بكامله.

كذلك عندما نقوم باختبار المصابيح المنتجة في أحد المعامل للوصول إلى معرفة متوسط عمر المصباح، لا يمكننا تجريب كافة المصابيح المنتجة ولكن نكتفي بتجريب عينة من المصابيح للوصول إلى القرار الذي يحدد متوسط عمر المصباح المنتج.

مما سبق نلاحظ أننا غالباً ما نقوم بدراسة العينة بدلاً من المجتمع لاستحالة أو صعوبة الدراسة للمجتمع بكامله حيث تؤدي دراسة العينة إلى التنبؤ بحال المجتمع، أو الوصول إلى قرار يتعلق بهذا المجتمع الإحصائي. ولكي يكون هذا القرار دقيقاً يجب اختيار العينة بشكل جيد، وهذا ما يدعى بعينة غير منحازة (unbiased) أي مختارة بشكل عشوائي.

5-3-1- العينة العشوائية:

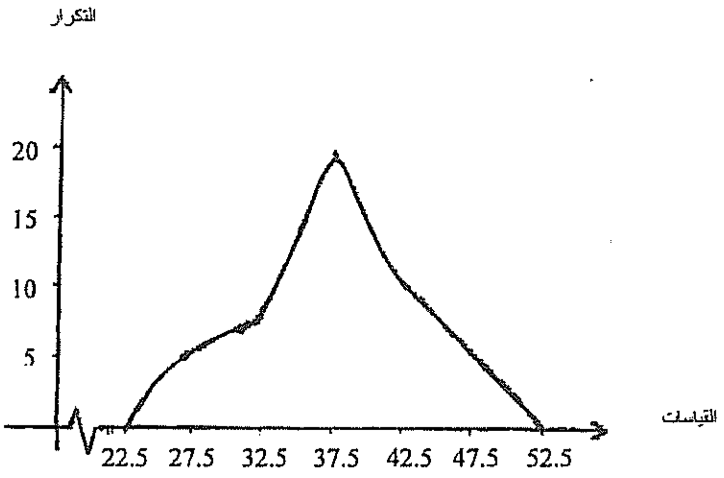
تعريف: العينة العشوائية هي عينة مختارة من المجتمع الإحصائي بحيث أن أي مجموعة جزئية أخرى من المجتمع لها نفس عدد عناصر هذه العينة، تملك الفرصة نفسها للاختيار.

فمثلاً لمعرفة نسبة المهندسين المدخنين في القطر السوري والذي يعتبر مجتمعاً إحصائياً

محدوداً نجد أن العينة التي تشتمل على مهندسين سوريين من كافة فروع الهندسة ومن الجنسين عينة جيدة أي عينة عشوائية ودراستها تعطينا نسبة جيدة وقريبة من الحقيقة بل يمكن القول أنها النسبة الحقيقية. وتجدر الإشارة إلى أنّ هنالك طرقاً متعددة لاختيار العينات العشوائية.

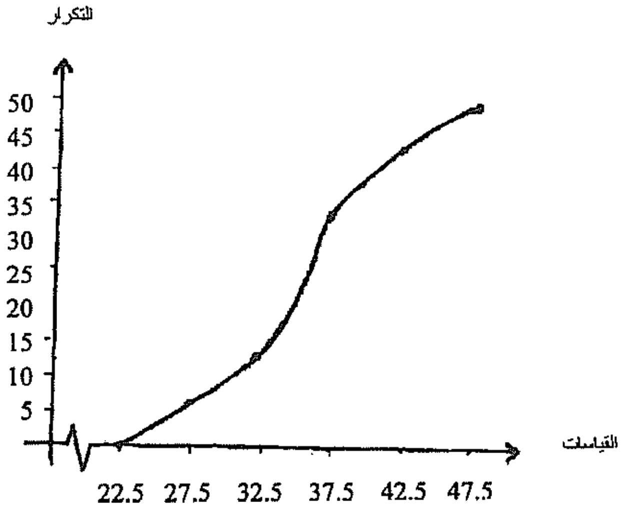
6-3-1- المنحني التكراري (The frequency curve):

المنحني التكراري هو الخط المنحني الذي يصل بين النقاط التي فواصلها مراكز الفئات وتراتيبها تكرارات هذه الفئات، وهو المضلع التكراري نفسه عندما يكون عدد الفئات كبيراً وطول كل فئة صغير حيث يتحول الخط المنكسر في هذه الحالة إلى خط منحني ندعوه بالمنحني التكراري. من التعريف السابق نلاحظ أن المنحني التكراري للبيانات المبينة في الجدول (2-1) هو كما في الشكل التالي:

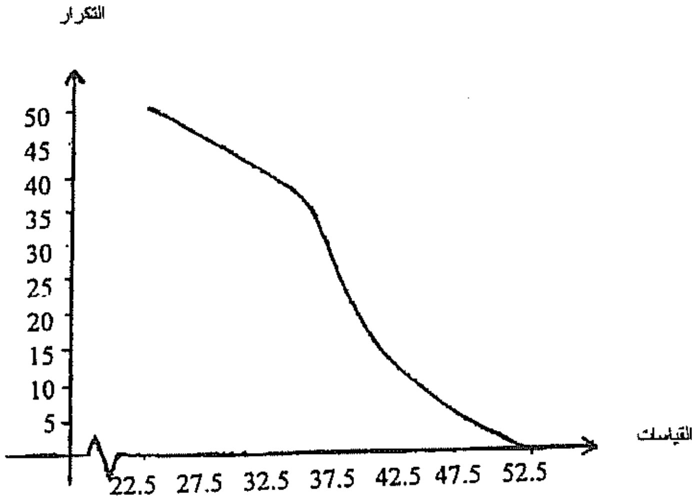


الشكل (6-1): المنحني التكراري

كما يمكننا تمثيل كلاً من المنحني التكراري المجمع الصاعد والمنحني التكراري المجمع النازل، لمثالنا السابق بالأشكال التالية:



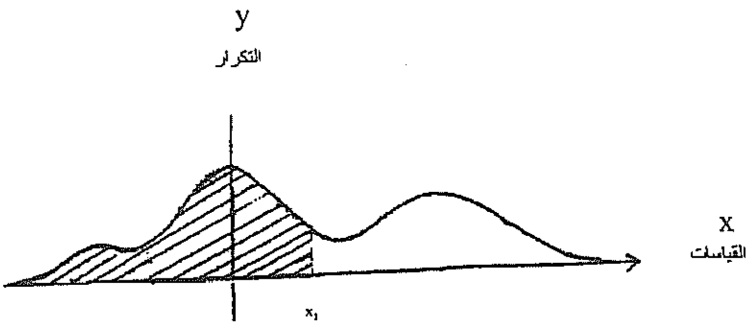
الشكل (1-7): المنحني التكراري المجمع الصاعد



الشكل (1-8): المنحني التكراري المجمع النازل

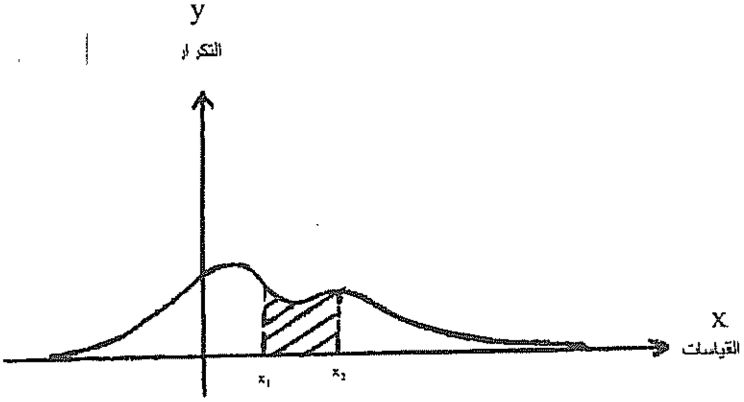
ممّا سبق وجدنا أنه يمكن تمثيل البيانات في أي مثالٍ إحصائي بمدرج تكراري على شكل مستطيلات وبالتالي بمضلع تكراري أو منحني تكراري. حيث تختلف أشكال المنحنيات التكرارية من مثال لآخر.

إن المساحة التي تقع تحت المنحني التكراري تتناسب مع التكرار الكلي للمجتمع، حيث نعتبر هذه المساحة عادةً (في الإحصاء) مساوية للواحد (واحدة المساحات) لجميع المنحنيات التكرارية. وبالتالي فإن المساحة المظللة في الشكل (1-9) مثلاً تمثل النسبة المئوية لعدد القياسات التي هي أصغر من x_1. بينما المساحة غير المظللة فهي تمثل النسبة المئوية لعدد القياسات الأكبر من x_1.



الشكل (1-9)

كما تمثل المساحة المظللة في الشكل (1-10) النسبة المئوية لعدد القياسات الواقعة بين x_1 و x_2.



الشكل (1-10)

1-4- مقاييس النزعة المركزية والتشتت:

1-4-1- الوسطاء والإحصائيات (Parameters and Statistics):

الوسيط: هو أيّ قيمة عددية تصف مجتمعاً إحصائياً. بينما الإحصائية هي أي قيمة عددية تصف عينة. فإذا كان لدينا مجموعة من البيانات تمثل مجتمعاً محدوداً فإن المتوسط الحسابي لهذه البيانات يكون وسيطاً وفي الحالة التي تمثل فيه البيانات عينة من مجتمع غير محدود، يكون المتوسط الحسابي لهذه البيانات إحصائية.

يُرمز لوسطاء المجتمع الإحصائي عادةً بأحرف يونانية، بينما يُرمز لإحصائيات العينة بحروف لاتينية، فمثلاً سنرمز للمتوسط الحسابي لمجتمع إحصائي ما بالرمز μ، بينما سنرمز للمتوسط الحسابي لعينة ما بالرمز $\bar{x}$.

من التعريف السابق نلاحظ أن الوسيط في المجتمع الإحصائي هو قيمة ثابتة تصف هذا المجتمع. بينما الإحصائية هي قيمة متغيرة تتغير من عينة لأخرى.

في حالة المجتمع الإحصائي المحدود يتم حساب الوسطاء لهذا المجتمع بشكل مباشر، أمّا في حالة المجتمع الإحصائي غير المحدود فيتم حساب الإحصائيات لعينة عشوائية من هذا المجتمع حيث تُؤخذ هذه الإحصائيات كقيم تقريبية لوسطاء المجتمع، أي سنستخدم إحصائيات العينة كتقدير لوسطاء المجتمع.

ولكن هل الإحصائيات التي نحصل عليها لعينة ما تعبر تعبيراً دقيقاً عن وسطاء المجتمع؟ وكم هي درجة هذه الدقة؟ وهل نكتفي بإحصائيات عينة واحدة أم نلجأ لاختيار عدة عينات من المجتمع؟. في الحقيقة سنلجأ لاختيار عدة عينات وسندرس توزيع الإحصائيات في هذه العينات المتعددة وسنحدد درجة الثقة بتقديرنا هذا. وهو ما سندرسه في فصول قادمة.

1-4-2- مقاييس النزعة المركزية (Measures of central tendency):

مقياس النزعة المركزية هو أي قياس عددي يعبر عن موضع تمركز التوزيع، لمجموعة من البيانات. ومن هذه المقاييس سندرس المتوسط الحسابي والوسط والمنوال.

(1) المتوسط الحسابي (The mean):

المتوسط الحسابي النقطي لمجموعة من القياسات $x_1, x_2, \ldots, x_n$ (ليست بالضرورة كلها مختلفة) الممثلة لمجمعٍ إحصائي محدود حجمه n، يُعرف بمجموع هذه القياسات مقسوماً على عددها n. أي أنَّ:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1.1)$$

أمّا إذا كانت القياسات $x_1, x_2, \ldots, x_n$ تمثل عينة حجمها n مأخوذة من مجتمع غير محدود فإن المتوسط الحسابي النقطي لهذه العينة يكون:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1.2)$$

هذا ويمكن كتابة Σx عوضاً عن $\sum_{i=1}^{n} x_i$ عندما لا يكون هناك أي التباس.

مثال (2): عند البحث في مسألة متوسط عمر المحرك لنوع معيّن من السيارات، أُخذت عينة مؤلفة من 9 سيارات من هذا النوع وخضعت محركاتها للتجريب ثم سُجلت أعمار هذه المحركات (بآلاف الكيلومترات) فكانت كما يلي:

140, 180, 175, 210, 157, 184, 140, 165, 180

أوجد متوسط عمر المحرك لهذا النوع من السيارات.

الحل:

$$\overline{x} = \frac{\Sigma x}{n} = \frac{140 + 180 + \ldots + 165 + 180}{9} = 170.11 \approx 170$$

أي أن متوسط عمر المحرّك لهذا النوع من السيارات هو 170 ألف كيلومتر تقريباً.

(2) حساب المتوسط الحسابي في حالة البيانات المبوبة:

إذا كانت القياسات المفروضة معرفة ضمن جدول التوزيع التكراري النقطي التالي:

القياسات	x_1	$x_2 \ldots$	x_k
التكرار	f_1	$f_2 \ldots$	f_k

حيث تمثل f_i التكرار للقياس x_i $(i = 1, 2, \ldots, k)$، عندئذٍ يعطى المتوسط الحسابي النقطي لهذه القياسات بالعلاقة:

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} \qquad (1.3)$$

وفي الحالة التي تكون فيها القياسات معرفة ضمن جدول التوزيع التكراري للبيانات المبوبة، يتم حساب المتوسط الحسابي* من العلاقة التالية:

$$\bar{x} = \frac{\sum_{i=1}^{k} c_i f_i}{\sum_{i=1}^{k} f_i} \qquad (1.4)$$

حيث يمثل k عدد الفئات كما تمثل c_i و f_i على الترتيب مركز الفئة i وتكرار هذه الفئة.

مثال (3): لنأخذ القياسات المبينة في المثال (1) والمطلوب حساب المتوسط الحسابي لهذه القياسات باستخدام (1.3) ثم (1.4).

الحل: لإيجاد المتوسط الحسابي النقطي لهذه العينة، نجد من العلاقة (1.3) أنّ:

$$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{(25)(3)+(28)(3)+...+(49)(2)}{50} = \frac{1900}{50} = 38$$

كذلك نجد من العلاقة (1.4) أن المتوسط الحسابي للبيانات المبوبة لهذه العينة هو:

$$\bar{x} = \frac{\Sigma cf}{\Sigma f} = \frac{(27.5)(6)+...+(47.5)(6)}{50} = \frac{1885}{50} = 37.7$$

ممّا سبق نلاحظ أن المتوسط الحسابي النقطي لهذه القياسات يعطي القيمة الصحيحة تماماً للمتوسط بينما يعطي المتوسط الحسابي للبيانات المبوبة قيمة تقريبية لهذا المتوسط وذلك لأننا قد نفقد أو نحصل على بعض الأعداد عند توزيع القياسات على فئات. ولكننا نلاحظ أيضاً أن الجوابين متقاربان نوعاً ما، ممّا يدل على أننا وفقنا في اختيار الفئات.

(3) الوسط (The median):

الوسط النقطي لمجموعة من القياسات $x_1, x_2, ..., x_n$ المرتبة تصاعدياً أو تنازلياً، هو القيمة الواقعة في وسط هذه البيانات (إذا كان عددها فردياً) أو المتوسط الحسابي للقيمتين الواقعتين في الوسط (إذا كان عدد القياسات زوجياً). ويرمز عادةً للوسط في حالة المجتمع الإحصائي بالرمز $\tilde{\mu}$ وفي حالة العينة بالرمز $\tilde{x}$. أمّا الوسط لمجموعة من البيانات المبوبة فهو القياس التقريبي الواقع في فئة الوسط والذي نحصل عليه من العلاقة:

---

* يصطلح على تسمية المتوسط الحسابي للبيانات المبوبة إلى فئات أسم المتوسط الحسابي الفئوي. وبالمثل يصطلح على تسمية المقاييس المتعلقة بالجداول المبوبة لفئات أسم المقاييس الفئوية.

$$\tilde{x} = L_k + \frac{\frac{N}{2} - \sum_{i=1}^{k-1} f_i}{f_k} \ell \qquad (1.5)$$

حيث N هو العدد الكلي للقياسات، k هو تسلسل فئة الوسط، و ℓ طول هذه الفئة أمّ L_k و f_k فهما على الترتيب الحد الأدنى والتكرار لفئة الوسيط.

ملاحظة ②: يتم الحصول على فئة الوسط من خلال البحث عن الفئة التي تحوي القياس الذي تسلسله $\frac{N}{2}$ بالاستفادة من تكرارات الفئات.

مثال (4): أوجد الوسط النقطي للعينة المعرفة في المثال (2).

الحل: نلاحظ أن عدد القياسات في المثال (2) هو عدد فردي يساوي 9، عندئذٍ نرتب القياسات ترتيباً تصاعدياً ثم نأخذ القيمة الواقعة في الوسط كما يلي:

140, 140, 157, 165,175, 180, 180, 184, 210

وبالتالي فإن الوسط النقطي لهذه العينة يكون $\tilde{x} = 175$

مثال (5): أوجد الوسط النقطي لمجموعة القياسات المذكورة في المثال (1)، وكذلك الوسط للبيانات المبوبة.

الحل: نلاحظ أن عدد القياسات في المثال (1) يساوي 50 وهو عدد زوجي، عندئذٍ وبعد ترتيب البيانات تصاعدياً نلاحظ أن القيمتين الواقعتين في الوسط هما 38 ، 38 وبالتالي فإن

الوسط النقطي لهذه العينة يكون: $\tilde{x} = \frac{38+38}{2} = 38$

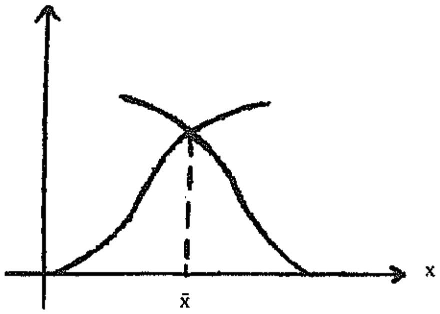
ولإيجاد الوسط للبيانات المبوبة نلاحظ من الجدول (1.2) أن فئة الوسط هي الفئة الثالثة لأن القياس الذي ترتيبه $\frac{N}{2} = \frac{50}{2} = 25$ يقع فيها، وبالتالي فإن قيمة الوسط تكون:

$$\tilde{x} = 35 + \frac{25-14}{20}(5) = 37.75$$

ملاحظة ③: لاحظ بأن الوسط النقطي يعطي القيمة الصحيحة تماماً، بينما الوسط للبيانات المبوبة فهو يعطي قيمة وهمية تقريبية.

ملاحظة ④: ليس بالضرورة أن يكون الوسط هو نفسه المتوسط الحسابي.

ملاحظة ⑤: يمكن الاستفادة من مفهومي التوزيع التكـراري المجمـع الصـاعد والتوزيـع التكراري المجمع النازل في تعيين وسط التوزيع التكراري أو وسط البيانات بشـكل هندسـي وذلك عن طريق نقطة تقاطع المنحنيين التكراريين الممثلين لهذين التوزيعين. كما في الشـكل التالي:



الشكل (1.11)

4- المنوال (The mode):

المنوال النقطي لمجموعة من القياسات هو تلك القيمة الأكثر تكراراً بينها ويُرمز لـه بـالرمز M.

من التعريف السابق نلاحظ أن المنوال قد لا يكون موجوداً، وإذا وجد فهو لـيس بالضـرورة وحيداً. أمّا المنوال لمجموعة من البيانات المبوبة فهو القياس الواقع في الفئة الأعلى تكـراراً، ويحسب من العلاقة التالية:

$$M = L_k + \frac{f_k - f_{k-1}}{(f_k - f_{k-1}) + (f_k - f_{k+1})}\ell \qquad \textbf{(1.6)}$$

حيث يمثل k تسلسل فئة المنوال، و ℓ طول هذه الفئة، أمّـا f_k، f_{k-1}، f_{k+1} فهـي تمثـل تكرارات كل من فئة المنوال والفئتان التي تسبقها وتليها على الترتيب.

مثال (6): أوجد المنوال النقطي للبيانات المذكورة في المثال (2).

الحل: نلاحظ أن كلاً من القيمتين 140 و 180 تكرر ذكرها مرتين في البيانات، وبالتالي فإنّ كلاً منهما يعتبر المنوال النقطي.

مثال (7): أوجد المنوال النقطي للبيانات المبينة في المثال (1)، وكذلك المنوال للبيانات المبوبة.

الحل: بالعودة للمثال (1) ومن الجدول (1-1) نلاحظ أنّ القياس 43 هو الأكثر تكراراً بين القياسات المذكورة حيث تكرر العدد 43 عشر مرّات وبالتالي المنوال النقطي لهذه البيانات هو 43 أي أنّ: M=43. ومن الجدول (1-2) نجد أن الفئة الثالثة هي الفئة الأعلى تكراراً حيث نلاحظ أن تكرارها هو 20 وبالتالي نجد أنّ المنوال للبيانات المبوبة هو:

$$M = 35 + \frac{(20-8)(5)}{(20-8)+(20-10)} = 37.72$$

ملاحظة ⑥: ممّا سبق نلاحظ بأنّ المنوال النقطي يعطي القيمة الصحيحة تماماً للمنوال بينما يعطي المنوال للبيانات المبوبة قيمة تقريبية وهمية لهذا المنوال.

ملاحظة ⑦: هناك علاقة تجريبية تربط بين المقاييس المركزية الثلاثة السابقة (المتوسط الحسابي والوسط والمنوال) سنذكرها في فقرة مقاييس الالتواء.

1-4-3- مقاييس التشتت (Measures of variations):

إن مقاييس النزعة المركزية الثلاثة المذكورة في الفقرة السابقة لا تعطي وحدها وصفاً دقيقاً لتوزيع البيانات. حيث يلزمنا أيضاً معرفة شكل انتشار هذه البيانات حول المتوسط، فقد يكون لدينا مجموعتان من البيانات لهما نفس المتوسط الحسابي ولكنهما مختلفتان في تشتت كل منهما عن ذلك المتوسط. ومن مقاييس التشتت سندرس المدى والانحراف المتوسط المطلق والتباين والانحراف المعياري وأخيراً الأرباع والمئينات.

(1) المدى (The range):

إن المدى النقطي R لمجموعة من القياسات هو الفرق بين الحدين الأعلى والأدنى لهـا. أمّـا المدى للقياسات المبوبة فهو الفرق بين مركزي الفئتين الأولى والأخيرة.

مثال (8): أوجد المدى النقطي للبيانات المذكورة في المثال (1)، وكـذلك المـدى للبيانـات المبوبة.

الحل: نلاحظ أنّ المدى النقطي في المثال (1) يساوي: 49-25=24 ، أمّـا المـدى للبيانـات المبوبة فهو يساوي: 47.5-27.5=20

إنّ المدى مقياس غير مُرضي تماماً للتشتت، خصوصاً إذا كان حجم العينة كبيـراً. ذلـك لأن المدى يعطي فقط الفرق بين أكبر قياس وأصغر قياس، ولا يخبرنا شيئاً عن توزيع القياسـات بينهما.

(2) الأرباع (Quartiles):

وهي القياسات التي تقسم البيانات إلى أربعة أقسام متسـاوية. فـإذا كانـت $x_1, x_2, \ldots, x_n$ مجموعة من القياسات المرتبة تصاعدياً، عندئذٍ ندعو بالربيع الأدنى القياس الذي يكون تسلسله (ترتيبه) عند ربع القياسات أي الذي يقع قبله 25% من البيانات، كما ندعو بالربيع الأعلى ذلك القياس الذي يكون تسلسله عند ثلاثة أرباع القياسات أي الذي يقع قبله 75% من البيانات. أمّا الربيع الثاني فهو القياس الذي يكون تسلسله عند منتصف البيانات، أيّ الذي يقع قبلـه 50% من البيانات.

من التعريف السابق نجد أنّ الربيع الثاني لمجموعة من البيانات يساوي وسطها.

يُرمز عادةً للأرباع الثلاث الأدنى والثاني والأعلى بـالرموز Q_1، Q_2، Q_3 علـى الترتيـب، ويمكن توضيح ذلك بالشكل التالي:

	0%	25%	50%	75%	100%	
		Q_1	Q_2	Q_3		← القياسات

إذا كان عدد القياسات n فردياً، فإن رتبة الربيع الأدنى تكون $\frac{n+1}{4}$ بينما تكون رتبة الربيـع الأعلى $\frac{3n+1}{4}$ أمّا الوسيط أو الربيع الثاني فتكون رتبته $\frac{2n+1}{4}$. وفي الحالة التي يكـون

فيها عدد القياسات n زوجياً ويقبل القسمة على 4 يتم حساب الربيعات الأدنى والأعلى بحساب متوسط العددين الموافقين لترتيب كل ربيع.

(3) نصف المدى الربيعي:

إذا كان لدينا مجموعة من القياسات المرتبة تصاعدياً، وكان Q_1 و Q_3 الربيعين الأدنى والأعلى لهذه القياسات على الترتيب عندئذ نسمي المقدار:

$$Q = \frac{Q_3 - Q_1}{2} \qquad (1.7)$$

بنصف المدى الربيعي النقطي. حيث يستفاد من هذا القياس في عملية استثناء أو حذف القيم المتطرفة كثيراً في البيانات. أمّا نصف المدى الربيعي لمجموعة من البيانات المبوبة، فهو المقدار الناتج من العلاقة (1.7) مع ملاحظة أن Q_1 و Q_3 في هذه الحالة يمثلان الربيعين الأدنى والأعلى للبيانات المبوبة على الترتيب. ويتم حساب Q_1 و Q_3 من العلاقات التالية:

$$Q_1 = L_k + \frac{\frac{n}{4} - \sum_{i=1}^{k-1} f_i}{f_k} . \ell \qquad (1.8)$$

$$Q_3 = L_k + \frac{\frac{3n}{4} - \sum_{i=1}^{k-1} f_i}{f_k} . \ell \qquad (1.9)$$

حيث تمثل k و ℓ و f_k و L_k في العلاقة (1.8) ترتيب فئة الربيع الأدنى وطولها وتكرارها والحد الأدنى فيها، بينما تمثل هذه الرموز في العلاقة (1.9) الترتيب لفئة الربيع الأعلى وطولها وتكرارها والحد الأدنى فيها.

مثال (9): أوجد الربيعين الأدنى والأعلى للبيانات المبوبة المذكورة في المثال (1) ثم استنتج نصف المدى الربيعي لهذه البيانات.

الحل: من الجدول (2-1) نلاحظ أن فئة الربيع الأدنى هي الفئة الثانية وفئة الربيع الأعلى هي الفئة الرابعة وبالتالي فإنّ:

$$Q_1 = 30 + \frac{12.5 - 6}{8}(5) = 34.06$$

$$Q_3 = 40 + \frac{37.5 - 34}{10}(5) = 41.75$$

وبالتالي فإن نصف المدى الربيعي للبيانات المبوبة لهذه القياسات هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{41.75 - 34.06}{2} = 3.845$$

مثال (10): أوجد الأرباع الثلاث ونصف المدى الربيعي النقطي لكلٍ من القياسات في الحالتين التاليتين:

أ - 67,59,50,74,72,69,70,65,58

ب- 55,65,72,71,70,69,67,58

الحل: لترتيب القياسات تصاعدياً كما يلي: 50,58,59,65,67,69,70,72,74

حيث نلاحظ أنّ عدد القياسات (n=9) فردياً وبالتالي فإن ترتيب الربيعي الأدنى Q_1 هو $\frac{n+1}{4} = 2.5$ أي القياس الثالث وترتيب الربيعي الأعلى Q_3 هو $\frac{3n+1}{4} = 7$ أي القياس السابع، أمّا ترتيب الربيع الثاني فهو $\frac{2n+1}{4} = 4.75$ أيّ القياس الخامس. وبالتالي $Q_1 = 59, Q_2 = 67, Q_3 = 70$ ويكون نصف المدى الربيعي النقطي لهذه القياسات:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{70 - 59}{2} = 5.5$$

كما يلاحظ في الحسابات بأن الوسط النقطي لهذه البيانات هو: $\tilde{\mu} = 67$

ب- بعد ترتيب القياسات ترتيباً تصاعدياً نجد أنّ: 55,58,65,67,69,70,71,72

حيث نلاحظ أنّ عدد القياسات (n=8) زوجياً وبالتالي فإن:

$$Q_2 = \frac{67 + 69}{2} = 68 \quad , \quad Q_3 = \frac{70 + 71}{2} = 70.5 \quad ,$$

$$Q_1 = \frac{58 + 65}{2} = 61.5$$

وبالتالي فإن نصف المدى الربيعي النقطي يكون:

$$Q = \frac{Q_3 - Q_1}{2} = 4.5$$

(4) المئينات والعُشيرات (Percentiles and Deciles):

المئين (المئوي) P_i لمجموعة من القياسات المرتبة تصاعدياً، هو القياس الذي يتجاوز في ترتيبه i٪ من القياسات ويقّل عن (i-1)٪ من القياسات الباقية (حيث: i = 1,2,...,99) فمن أجل أي توزيع يوجد (99) مئين ممكن وهي: $P_1, P_2, ..., P_{99}$. فمثلاً P_{39} يدّل على القياس الذي يتجاوز في ترتيبه 39٪ من القياسات. وبالتالي فإن الربيع الأدنى هو P_{25} والوسط هو P_{50} بينما الربيع الأعلى فهو P_{75}.

بنفس الطريقة نجد أن العُشير (العشري) D_i لمجموعة من القياسات المرتبة تصاعدياً هو القياس الذي يتجاوز في ترتيبه $\frac{i}{10}$ من القياسات. فمن أجل أي توزيع يوجد (9) عُشيرات هي: $D_1, D_2, ..., D_9$ وبالتالي فإن D_2 هو القياس الذي يتجاوز في ترتيبه $\frac{2}{10}$ أو 20٪ من القياسات.

(5) تعيين الأرباع والمئينات والعُشيرات عن طريق المئينات:

لاحظنا أنه يمكن تعيين الأرباع والعُشيرات عن طريق المئينات (المئويات). لذلك سنبحث في طريقة إيجاد المئوي. فإذا كان لدينا مجموعة من القياسات عددها n فإن ترتيب المئوي i يكون $n\frac{i}{100}$، حيث نجد أنه بالإمكان حساب المئويات من خلال جدول التوزيع التكراري المجمع الصاعد، وبافتراض أنّ القياسات في فئة ما موزعة بانتظام بين طرفي المجال وذلك كما في المثال التالي:

مثال (11): الجدول التالي يمثل توزيع الرواتب لـ 140 موظف في إحدى الشركات الخاصة (بآلاف الليرات السورية):

الفئة	7<9	9<11	11<13	13<15	15<17	17<19	19<21	21<23
التكرار	8	12	25	40	20	17	13	5

والمطلوب حساب المئوي P_{35} والربيع الأعلى والعُشير D_3.

الحل: نشكل أولاً جدول التوزيع التكراري المجمع الصاعد كما يلي:

الجدول (5-1)

التكرار المجمع الصاعد	الحدود الدنيا للفئات
0	أقل من 7
8	أقل من 9
20	أقل من 11
35	أقل من 13
85	أقل من 15
105	أقل من 17
122	أقل من 19
135	أقل من 21
140	أقل من 23

إن ترتيب المئوي P_{35} هو: $140\times\frac{35}{100}=49$ أي أنّ المئوي P_{35} يقع في الفئة الرابعة ضمن جدول التوزيع التكراري الفئوي، أيّ أنّه أكبر من 13 وأصغر من 15 وترتيبه الرابع بين عناصر هذه الفئة (لأن مجموع عناصر الفئة الأولى والثانية والثالثة 45). أي أنّ قيمته:

$$P_{35}=13+(15-13).\frac{4}{40}=13.2$$

أيّ المئوي P_{35} هو الموافق للراتب 13200 ل.س. وبالتالي فإنّ 35% من الموظفين في هذه الشركة يتقاضون أقل من 13200 ليرة سورية شهرياً.

كذلك نلاحظ أن ترتيب الربيع الأعلى $Q_3=P_{75}$ هو: $140\times\frac{75}{100}=105$ وبالتالي فهو ينطبق على الحد الأدنى للفئة السادسة في جدول التوزيع التكراري الفئوي. أيّ أنّ: $Q_3=P_{75}=17$ (وذلك بملاحظة جدول التوزيع التكراري المجمع الصاعد). وبالتالي فإنّ 75٪ من الموظفين يتقاضون أقل من 17000 ل.س.

وأخيراً نلاحظ أن ترتيب العُشير الثالث $D_3=P_{30}$ هو: $140\times\frac{30}{100}=22$ فهو ينتمي إلى الفئة الثالثة من جدول التوزيع التكراري الفئوي وبما أنّ مجموع عناصر الفئتين الأولى والثانية هو 20 فترتيبه في الفئة الثالثة هو الثاني والعشرون أي أن قيمته هي:

$$D_3 = P_{30} = 11 + (13-11)\frac{22}{25} = 12.76$$

أيّ أن العُشير D_3 هو الموافق للراتب 12760. وبالتالي فإنّ 30٪ من الموظفين رواتبهم أقل من 12760 ل.س.

ممّا سبق نلاحظ أنّ الأرباع والعُشيرات والمئويات هي أكثر فعالية من المدى في التعبير عن تبعثر البيانات الإحصائية.

(6) الانحراف المتوسط المطلق (Mean absolute deviation):

الانحراف المتوسط المطلق النقطي لمجموعة من القياسات $x_1, x_2, \ldots, x_n$ والتي متوسطها $\bar{x}$، يُعطى بالعلاقة:

$$MD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} \qquad (1.10)$$

أيّ هو معدل القيم المطلقة لانحرافات هذه القياسات عن متوسطها. أمّا الإنحراف المتوسط المطلق لمجموعة من البيانات المبوبة فهو:

$$MD = \frac{\sum_{i=1}^{k} f_i |c_i - \bar{x}|}{n} \qquad (1.11)$$

حيث يمثل k هنا عدد الفئات أمّا c_i فهي مراكز الفئات.

(7) التباين (The variance):

إنّ التباين النقطي لمجموعة من القياسات $x_1, x_2, \ldots, x_n$ الممثلة لعينة ما والتي متوسطها $\bar{x}$ هو معدل مربعات انحرافات هذه القياسات عن متوسطها ونرمز له بـ S^2 حيث:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \qquad (1.12)$$

إنّ العلاقة (1.12) تعطي تباين العينة وهو مستخدم كتقدير لتباين المجتمع، وقد وجد إحصائياً بأن هذا التقدير يكون أفضل عند القسمة على n-1 بدلاً من n في تلك العلاقة حيث تصبح بالشكل:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \qquad (1.13)$$

ويمكن حساب تباين العينة أيضاً من العلاقة التالية المكافئة للعلاقة (1.13):

$$S^2 = \frac{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}{n(n-1)} \qquad (1.14)$$

أمّا التباين لمجموعة عددها n من البيانات المبوبة فهو:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(c_i - \overline{x})^2 f_i \qquad (1.15)$$

حيث يمثل k عدد الفئات أمّا c_i فهي تمثل مركز الفئة رقم i .

ملاحظة ⑧: يرمز عادةً لتباين العينة بالرمز S^2 كما يرمز لتباين المجتمع بالرمز σ^2 وهو غالباً ما يكون غير معلوم.

(8) الانحراف المعياري (Standard deviation):

الانحراف المعياري النقطي لمجموعة من القياسات $x_1, x_2, \ldots, x_n$ هو الجذر التربيعي الموجب لتباينها النقطي.

أمّا الانحراف المعياري لمجموعة مبوبة من البيانات فهو الجذر التربيعي الموجب لتباينها.

مثال (12): أوجد التباين والانحراف المعياري للعينة المؤلفة من القياسات المذكورة في المثال (2).

الحل: وجدنا في المثال (2) أنّ المتوسط الحسابي لهذه العينة هو: $\overline{x} = 170.11$. وباستخدام العلاقة (1.13) نجد أن تباين العينة هو:

$$S^2 = \frac{\sum(x-\overline{x})^2}{n-1}$$

$$= \frac{(140-170.11)^2 + (180-170.11)^2 + \ldots + (180-170.11)^2}{8} = 501.86$$

وبالتالي فإن الانحراف المعياري يكون:

$$S = \sqrt{S^2} = \sqrt{501.86} = 22.4$$

ملاحظة ⑨: عندما يكون عدد القياسات كبيراً يُفضل استخدام العلاقة (1.14) لحساب التباين وذلك لأن العلاقة (1.13) تصبح أكثر تعقيداً في هذه الحالة.

ملاحظة (10): يمكن باستخدام الآلة الحاسبة الحصول على التباين النقطي بشكل سريع دون الرجوع للعلاقتين (1.13) و (1.14) السابقتين.

1-4-4- معامل الاختلاف (Coefficiont of Variation):

معامل الاختلاف لمجموعة من القياسات هو النسبة بين الانحراف المعياري والمتوسط الحسابي لهذه القياسات. ويرمز له بالرمز C.V حيث:

$$C.V = \frac{S}{\overline{X}} \qquad (1.16)$$

وفي حالة المجتمع المحدود يكون: $C.V = \frac{\sigma}{\mu}$

كما يُعرَّف معامل الاختلاف أيضاً بأنه وحدة الانحراف عن المتوسط الحسابي أو مقدار الانحراف في الوحدة الواحدة من البيانات.

ويمكن حساب معامل الاختلاف باستخدام قاعدة الربيعات من العلاقة التالية:

$$C.V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad (1.17)$$

يستخدم معامل الاختلاف للمقارنة بين مجموعتين من البيانات وتحديد أيٍّ من هاتين المجموعتين أفضل من الأخرى وذلك بناءً على مقدار وحدة الانحراف، فإذا تباين معاملي اختلاف مجموعتين فإن المجموعة الأفضل هي الأصغر معاملاً.

ملاحظة (11): إن معامل الاختلاف السابق يسمى معامل الاختلاف النقطي. أمّا معامل الاختلاف لمجموعة من القياسات المبوبة فهو النسبة بين الانحراف المعياري والمتوسط الحسابي لهذه القياسات.

مثال (13): قارن بين المجموعتين التاليتين من البيانات الممثلتين لعينتين مختلفتين من حيث قوة التمركز حول المتوسط الحسابي ودرجة التشتت.

A:	30	4	5	7	8	
B:	18	20	22	19	21	23

الحل: لنحسب المتوسط الحسابي والانحراف المعياري لكلا المجموعتين حيث نجد أنّ:

$\overline{x} = 10.8$ ، $S = 10.849$ من أجل البيانات A كما نجد أنّ:

$S = 1.871$ ، $\bar{x} = 20.50$ من أجل البيانات B

وبالتالي فإن معامل الاختلاف لكلٍ من هاتين المجموعتين يكون:

$$C.V_A = \frac{10.849}{10.8} = 1.005$$

$$C.V_B = \frac{1.871}{20.50} = 0.091$$

حيث نلاحظ أنّ مجموعة البيانات B أفضل من المجموعة A. لأن درجة التشتت في مفردات مجموعة البيانات A عالية وقوة التمركز حول المتوسط الحسابي ضعيفة بعكس مفردات المجموعة B التي نلاحظ أن درجة التشتت فيها ضعيفة وقوة التمركز لهـا حول المتوسط الحسابي أفضل من مفردات المجموعة A.

1-4-5- مقاييس الالتواء والتفلطح:

(1) مقاييس الالتواء: (معامل الالتواء المنوالي – معامل الالتواء الوسطي – معامل الالتواء الربيعي):

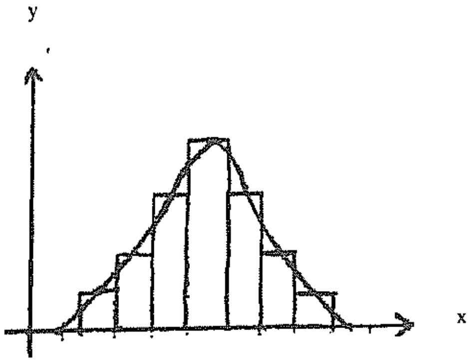
يُقال عن التوزيع التكراري بأنه متناظر (متماثل) أو معتدل (symmetric) إذا تساوى متوسطه الحسابي مع كل من منواله ووسيطه أي إذا كان: $\bar{x} = M = \tilde{x}$. وخلاف ذلك أي إذا اختلف المتوسط الحسابي عن كل من الوسيط والمنوال عندئذٍ يُقال بأن التوزيع التكراري غير متناظر (غير متماثل) أو ملتوٍ. حيث يكون الالتواء موجباً (نحو اليمين) إذا كان انتشار البيانات على يمين المنوال أكبر من انتشارها على يساره. وبعكس ذلك يكون الالتواء سالباً (نحو اليسار).

وهكذا نلاحظ أن الالتواء يقيس درجة تماثل أو اعتدال التوزيع التكراري حول متوسط البيانات وعلى هذا الأساس يُعرف معامل الالتواء المنوالي sk بالعلاقة التالية:

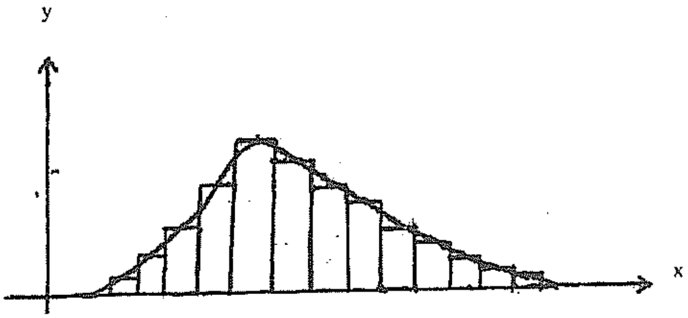
$$sk = \frac{\bar{x} - M}{s} \qquad \textbf{(1.18)}$$

من العلاقة السابقة نجد أن معامل الالتواء يساوي صفراً بالنسبة للتوزيعات المتماثلة (المتناظرة) ونجد أيضاً أنه كلما امتد المتوسط الحسابي باتجاه البيانات الكبيرة وبعيداً إلى يمين المنوال كان الالتواء كبيراً وموجباً وبعكس ذلك يكون الالتواء سالباً.

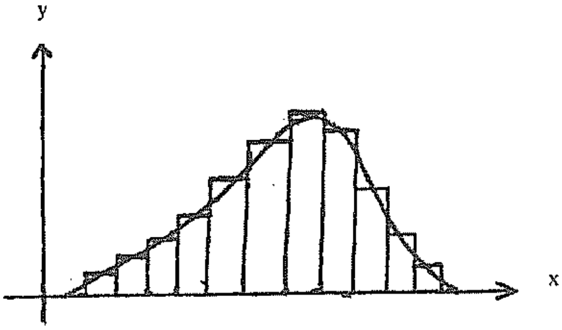
ويمكن توضيح ذلك من خلال الأشكال البيانية التالية:



الشكل (1-12): توزيع تكراري متماثل أو معتدل فيه $\bar{x} = M = \tilde{x}$ ودرجة الالتواء صفر



الشكل (1-13): توزيع تكراري غير متماثل وملتو نحو اليمين (موجب) فيه $M < \bar{x} < \tilde{x}$



الشكل (1-14): توزيع تكراري غير متماثل وملتو نحو اليسار (سالب) $\tilde{x} < \bar{x} < M$

ويمكن استخدام الوسط $\tilde{x}$ في حساب معامل الالتواء من خلال الدستور التالي:

$$sk = \frac{3(\bar{x} - \tilde{x})}{s} \qquad (1.19)$$

والذي يُدعى بمقياس بيرسون للالتواء أو معامل الالتواء الوسطي. حيث يعتبر مقياس بيرسون للالتواء أكثر دقة من معامل الالتواء المنوالي وذلك لأن الوسط يصف البيانات بشكلٍ أفضل من المنوال.

في الحالة التي يكون فيها حساب المتوسط الحسابي متعذراً كما في التوزيعات المفتوحة نقوم بحساب معامل الالتواء من العلاقة التالية:

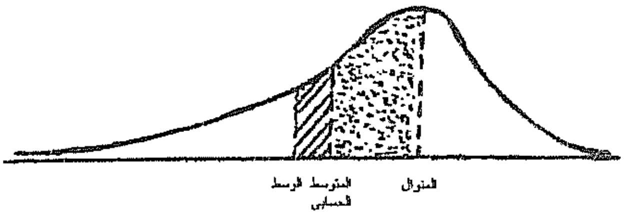
$$sk = \frac{(Q_3 - \tilde{x}) - (\tilde{x} - Q_1)}{Q_3 - Q_1} \qquad (1.20)$$

والتي تعطي معامل الالتواء الربيعي. حيث يكون هذا المعامل مساوياً للصفر عندما يكون التوزيع متماثلاً وذلك لأن الربيعين يكونان على بعدين متساويين من الوسط وعلى جانبيه. وفي الحالة التي يكون فيها التوزيع ملتوياً نحو اليمين نجد أنّ Q_3 تكون مبتعدة كثيراً عن الوسط مقارنة بالربيع الأدنى Q_1 ويكون حاصل الفرق في بسط العلاقة (1.20) موجباً وبالتالي يكون معامل الالتواء موجباً، وبالعكس عندما يكون الالتواء نحو اليسار نجد أنّ Q_1 تكون مبتعدة كثيراً عن الوسيط مقارنةً بالربيع الأعلى Q_3 ويكون معامل الالتواء في هذه الحالة سالباً.

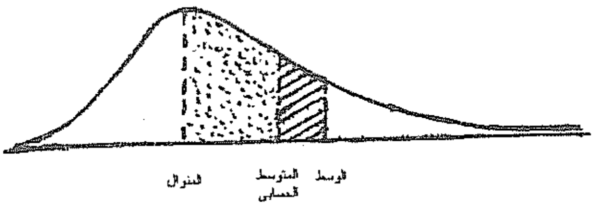
ملاحظة (12) (العلاقة التجريبية المركزية): هناك علاقة تربط بين مقاييس النزعة المركزية الثلاثة (المتوسط الحسابي والوسط والمنوال) حيث نجد أنّ:

$$(\bar{x} - M) = 3(\bar{x} - \tilde{x}) \qquad (1.21)$$

ويمكن توضيح ذلك من خلال الأشكال (1-15) و (1-16) التالية:



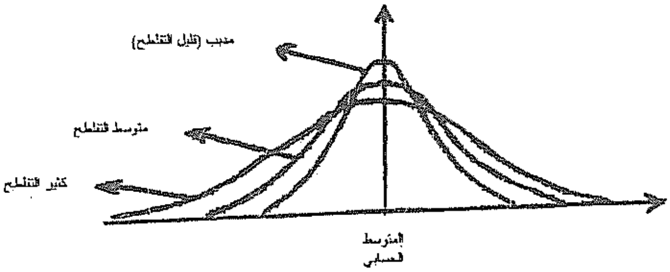
الشكل (1-15)



الشكل (1-16)

(2) مقاييس التفلطح (معامل التفلطح):

يُقال عن التوزيع التكراري للبيانات بأنه متفلطح (kurtosis) عندما يكون انتشار البيانات حول متوسطها متماثلاً وبعيداً، أيّ أنّ تشتت البيانات على جانبي المتوسط الحسابي يكون بصورة متماثلة (متناظرة) ومتطرفة. وهكذا نجد أنّ التفلطح يقيس ابتعاد البيانات أو اقترابها من متوسطها الحسابي ويمكن توضيح ذلك من خلال الشكل التالي:



الشكل (1.17)

يقاس معامل التفلطح من العلاقة التالية:

$$k=\frac{\sum(x_i-\overline{x})^4}{n\ \ s^4} \qquad (1.22)$$

مثال (14): لتكن البيانات التالية والتي تمثل عينة عشوائية مختارة:

30,4,5,7,8

والمطلوب حساب معاملي الالتواء الوسطي والتفلطح لهذه البيانات.

الحل: من البيانات السابقة نلاحظ أنّ:

$$\tilde{x} = 7$$

$$\bar{x} = \frac{\sum x}{n} = 10.8,\ S^2 = \frac{\sum(x-\bar{x})^2}{n-1} = 117.701,\ S = 10.849,\ S^4 = 13853.525$$

وبالتالي نجد أنّ معاملي الالتواء الوسطي والتفلطح يكونا:

$$sk = \frac{3(\bar{x}-\tilde{x})}{s} = 1.051$$

$$k = \frac{\sum(x_i-\bar{x})^4}{n\ s^4} = \frac{(30-10.8)^4+(4-10.8)^4+(5-10.8)^4+(7-10.8)^4+(8-10.8)^4}{69267.625}$$

$$= \frac{139435.2164}{69267.625} = 2.01$$

وهو ما يعني أنّ البيانات متشتتة عن بعضها بنسبة $sk = 1.051$ على الأقل، وما يُشير أيضـــاً إلى تفلطح التوزيع بدرجة عالية من خلال معامل التفلطح $k = 2.01$.

5-1- المتغير المعياري:

إذا كان لدينا مجموعة من القياسات متوسطها الحسابي $\bar{x}$ وانحرافها المعياري s عندئذٍ يُعرف المتغير المعياري (أو الدرجة المعيارية) z لقياس ما x منها بأنه النسبة بين انحراف القياس x عن المتوسط الحسابي إلى الانحراف المعياري لهذه القياسات. أيّ أنّ:

$$z = \frac{x-\bar{x}}{s} \qquad \textbf{(1.23)}$$

وفي حالة المجتمع الإحصائي المحدود الذي متوسطه μ وانحرافه المعياري σ تصبح الدرجة المعيارية z للقياس x بالشكل:

$$z = \frac{x-\mu}{\sigma} \qquad \textbf{(1.24)}$$

وتستخدم الدرجة المعيارية عادةً للمقارنة بين قياسين مـــأخوذين مـــن عينتـــين مختلفتـــين (أو مجتمعين إحصائيين محدودين مختلفين)

مثال (15): تقدم طالب في قسم هندسة الحواسيب إلى امتحاني الرياضيات والإحصاء فحصل على الدرجة 82 في الرياضيات والدرجة 89 في الإحصاء. فإذا علمت أنّ متوسط درجـــات الطلاب المتقدمون لامتحان الرياضيات كان 68 وبانحراف معياري قـــدره 8، وأن متوســـط درجات الطلاب في الإحصاء كان 80 وبانحراف معياري قدره 6. بيّن في أي امتحان كانـــت درجة الطالب أفضل مقارنةً مع زملائه.

الحل: نلاحظ من مثالنا أنّ لدينا مجتمعين إحصائيين محدودين مختلفين، الأول يمثل مجموعة درجات طلاب قسم هندسة الحواسيب في مقرر الرياضيات والثاني يمثل مجموعة درجات هؤلاء الطلاب في الإحصاء. والقياسان المأخوذان من هذين المجتمعين هما درجتا الطالب في الامتحانين.

عندئذٍ يمكننا كتابة ما يلي:

المقرر	درجة الطالب	المتوسط الحسابي	الانحراف المعياري
الرياضيات	$x_1 = 82$	$\mu_1 = 68$	$\sigma_1 = 8$
الإحصاء	$x_2 = 89$	$\mu_2 = 80$	$\sigma_2 = 6$

ولمعرفة في أيّ الامتحانين كان الطالب أفضل مقارنةً مع زملائه نوجد الدرجة المعيارية لكل من القياسين x_2, x_1 حيث نجد أنّ:

$$z_1 = \frac{x_1 - \mu_1}{\sigma_1} = \frac{82-68}{8} = 1.75$$

$$z_2 = \frac{x_2 - \mu_2}{\sigma_2} = \frac{89-80}{6} = 1.50$$

أي أنّ درجة الطالب في الرياضيات تزيد 1.75 انحرافاً معيارياً عن متوسط درجات الطلاب في الرياضيات بينما تزيد درجة الطالب في الإحصاء بمقدار 1.50 انحرافاً معيارياً عن متوسط درجات الطلاب في الإحصاء. وبمقارنة هاتين الدرجتين المعياريتين نجد أنّ الطالب كان أكثر تفوقاً في الرياضيات مقارنة مع زملائه.

1-6 - مبرهنة تشيبتشيف (Chebyshev's theorem):

يعتبر المتوسط الحسابي والانحراف المعياري لمجموعة من القياسات، من المقاييس الهامة في الإحصاء. والمبرهنة التالية تبين أنه يمكننا تحديد توزيع أيّ مجموعة من القياسات من خلال معرفتنا بالمتوسط الحسابي والانحراف المعياري لها.

لقد اكتشف العالم الروسي تشيبتشيف أن عدد القياسات التي تقع بين قيمتين متناظرتين حول المتوسط الحسابي يرتبط بقيمة الانحراف المعياري لها.

مبرهنة تشيبتشيف:

إذا كانت $x_1, x_2, \ldots, x_n$ مجموعة من القياسات، وكان k عدداً يساوي الواحد أو أكبر، فعندئذ يوجد على الأقل النسبة $(1-\frac{1}{k^2})$ من هذه القياسات ضمن الفترة المغلقة $[\overline{x}-ks, \overline{x}+ks]$ حيث $\overline{x}$ هو المتوسط الحسابي للعينة و s هو الانحراف المعياري لها.

وفي حالة مجتمع إحصائي محدود متوسطه μ وانحرافه المعياري σ، يمكننا تطبيق مبرهنة تشيبتشيف مع تبديل $\overline{x}$ بـ μ و s بـ σ أي أنّ الفترة المغلقة المذكورة بتلك المبرهنة تصبح بالشكل: $[\mu-k\sigma, \mu+k\sigma]$

البرهان:

لتكن $x_1, x_2, \ldots, x_n$ مجموعة من القياسات الممثلة لعينة عشوائية. عندئذ نجد أنّ تباين هذه العينة هو:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}$$

أي أنّ:

$$ns^2 = \sum_{i=1}^{n}(x_i-\overline{x})^2$$

لكن $\sum_{i=1}^{n}(x_i-\overline{x})^2$ هو مجموع مربعات كل الانحرافات، لكافة القياسات عن المتوسط $\overline{x}$.

إذا أخذنا بعين الاعتبار فقط انحرافات العناصر الواقعة خارج الفترة $[\overline{x}-ks, \overline{x}+ks]$ فإننا سنحصل على مجموع أقل وهو:

$$\sum(x_r-\overline{x})^2 \ , x_r \notin [\overline{x}-ks, \overline{x}+ks]$$

وبما أن النقاط x_r تقع خارج الفترة السابقة فإن:

$$(x_r-\overline{x})^2 \geq (ks)^2 = k^2s^2$$

بفرض أنّ عدد هذه القياسات الواقعة خارج الفترة السابقة هو P فإنّ:

$$x_r \notin [\overline{x}-ks, \overline{x}+ks] \text{ حيث } \sum(x_r-\overline{x})^2 \geq pk^2s^2$$

لكن: $$ns^2 = \sum_{i=1}^{n}(x_i-\overline{x})^2 \geq \sum(x_r-\overline{x})^2 \geq pk^2s^2$$

أيّ أنّ: $n \geq pk^2$ وبالتالي $p \leq \frac{n}{k^2}$

أي أنّ عدد القياسات الواقعة خارج الفترة: $[\bar{x}-ks, \bar{x}+ks]$ هو على الأكثر $\frac{n}{k^2}$ وهذا يعني أن عدد هذه القياسات يمثل على الأكثر النسبة $(\frac{1}{k^2})$ من عدد القياسات الكلية. وبالتالي نجد أنّ عدد القياسات الواقعة داخل الفترة $[\bar{x}-ks, \bar{x}+ks]$ هو:

$$n-p \geq n-\frac{n}{k^2}=n \left(1-\frac{1}{k^2}\right)$$

أي أنّ عدد هذه القياسات يمثل على الأقل النسبة $(1-\frac{1}{k^2})$ من عدد القياسات الكلي. وهو المطلوب.

مثال (16): أخذنا عينة مؤلفة من 40 طالباً وقيست أوزانهم، فتبيّن أن متوسط العينة $\bar{x} = 62$، وانحرافها المعياري $s = 4kg$، والمطلوب تعيين المجال الذي يحوي، على الأقل، على 30 طالباً من هذه العينة.

الحل: بتطبيق مبرهنة تشيبتشيف نجد أنّ $(1-\frac{1}{k^2})$ على الأقل من عدد القياسات الكلية تقع في الفترة $[\bar{x}-ks, \bar{x}+ks]$. ولتحديد هذه الفترة نحسب قيمة k حيث نجد أنّ:

$$1-\frac{1}{k^2}=\frac{30}{40}=0.75$$

ومنه نجد أنّ $k=2$ ، وبالتالي فإن 30 طالب على الأقل تقع أوزانهم في الفترة:

$$[\bar{x}-ks, \bar{x}+ks]= [62\text{-}2(4),\ 62\text{+}2(4)] = [54,70]$$

إذن يمكننا القول بأنه على الأقل 75% من الطلاب تقع أوزانهم بين 54 و 70 كيلوغرام.

ملاحظة (13): إنّ عبارة على الأقل في مبرهنة تشيبتشيف تضعف من النتائج التي نحصل عليها، فالمبرهنة متحفظة لأنها تطبق على أيّ توزيع للقياسات دون استثناء. حيث نلاحظ أنّ المبرهنة تعطينا الحد الأدنى فقط للنسبة العددية ولا تعطينا أيّ فكرة فيما إذا كانت النسبة تمثل أكثر من ذلك الحد.

مثال (17): تنتج إحدى الشركات مصابيح كهربائية. لدراسة الجودة لهذه المصابيح أُخذَت عينة عشوائية مؤلفة من 500 مصباح حيث وجِد بعد فحصها أنّ المصابيح تعمل وسطياً لمدّة 800 ساعة وبانحراف معياري قدره 40 ساعة. ما هو العدد الأعظمي للمصابيح التي ستعمل لمدة أقل من 700 ساعة.

الحل: لدينا عينة عشوائية حيث: $n = 500$ ، $\overline{x} = 800$ ، $s = 40$

النسبة العظمى للمصابيح التي ستعمل لمدة أقل من 700 ساعة = $1-$ النسبة الدنيا للمصابيح التي ستعمل لمدة أكبر أو تساوي 700 (وهي المصابيح الواقعة في الفترة المغلقة التي حدّها الأدنى 700).

وبالتالي وحسب مبرهنة تشيبيتشيف نجد أنّ: $\overline{x} - ks = 700$ ومنه نجد أنّ: $k = 2.5$

حيث تكون النسبة المئوية الدنيا للمصابيح التي ستعمل لمدّة أكبر أو تساوي 700 ساعة هي:

$1-\frac{1}{k^2}$ وبالتالي فإنّ النسبة المئوية العظمى للمصابيح التي ستعمل لمدة أقل من 700 ساعة هي:

$$1-\left(1-\frac{1}{k^2}\right)=\frac{1}{k^2}=0.16$$

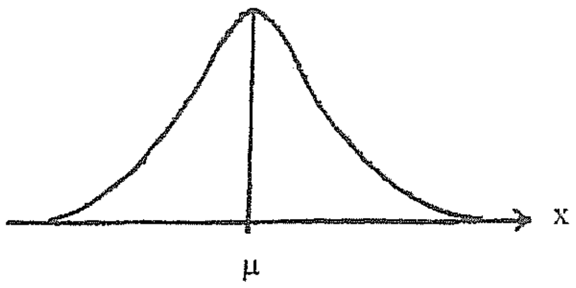
ومنه نجد أنّ العدد الأعظمي للمصابيح التي ستعمل لمدة أقل من 700 ساعة هو:

$$n\left(\frac{1}{k^2}\right)=(500)(0.16)=80$$

أي يمكننا القول بأنّ 80 مصباح على الأكثر من 500 مصباح ستعمل لمدة أقل من 700 ساعة وهي تمثل نسبة 16% على الأكثر من المصابيح المنتجة في هذه الشركة.

7-1- القاعدة التجريبية (Empirical rule):

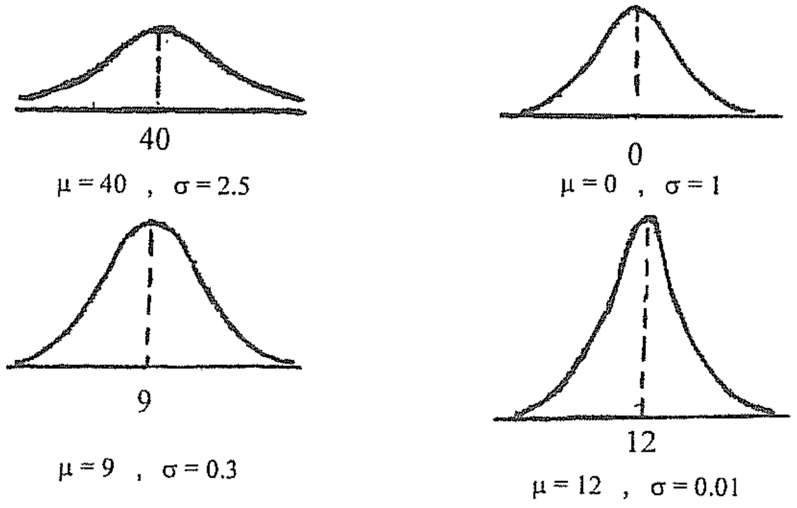
وجدنا سابقاً أنّ لكل توزيع تكراري مضلع تكراري أو بشكل أشمل منحني تكراري. حيث تختلف أشكال هذه المنحنيات التكرارية من توزيع لآخر. لكننا نلاحظ أن أغلب التوزيعات التي تواجهنا عملياً لها منحنيات تكرارية متشابهة، تأخذ شكل الجرس كما في الشكل التالي:



الشكل (1.18)

فمثلاً توزيع درجات امتحان مقررٍ ما لصف معين، يأخذ منحنيه شكل الجرس، كذلك نلاحظ أن قياسات أطوال طلاب الجامعة تتوزع بحيث يكون لها منحني تكراري له الشكل السابق أيضاً.

إنّ التوزيع السابق يدعى التوزيع الطبيعي، وهو من المفاهيم الهامة في الإحصاء. حيث يكون شكل المنحني لهذا التوزيع متناظراً بالنسبة للمتوسط الحسابي μ. ويختلف باختلاف قيمة الانحراف المعياري σ. فيما يلي بعض المنحنيات التكرارية لتوزيعات طبيعية توضح ذلك:



الشكل (1.19): بعض المنحنيات التكرارية الممثلة لتوزيعات طبيعية

وجدنا في مبرهنة تشيبتشيف أنه بإمكاننا معرفة النسب لتوزيع أيّ مجموعة من القياسات بمجرد معرفتنا بالمتوسط الحسابي والانحراف المعياري لهذه القياسات، حيث تطبق هذه المبرهنة على أي توزيع سواءً كان توزيع طبيعي أو غير طبيعي.

إذا كان التوزيع طبيعياً (أو قريب من الطبيعي، أيّ أن شكل منحنيه التكراري قريب من الجرس) عندئذٍ يمكننا تحديد النسب لتوزيع القياسات بشكلٍ أدق، وفق القاعدة التالية:

القاعدة التجريبية:

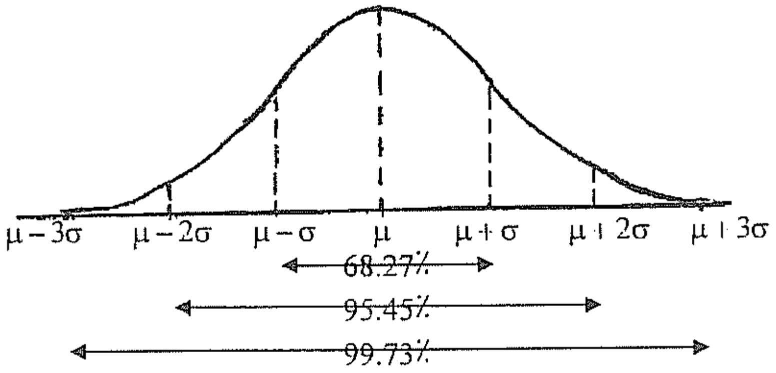
إذا كان لدينا مجموعة من القياسات لها التوزيع الطبيعي (أو القريب من الطبيعي)، فإنّه يكون لدينا تقريباً:

68.27% من القياسات تقع في المجال $[\mu - \sigma \;,\; \mu + \sigma]$

95.45% من القياسات تقع في المجال $[\mu - 2\sigma \;,\; \mu + 2\sigma]$

99.73% من القياسات تقع في المجال $[\mu - 3\sigma \;,\; \mu + 3\sigma]$

ويمكن توضيح ذلك بالشكل التالي:

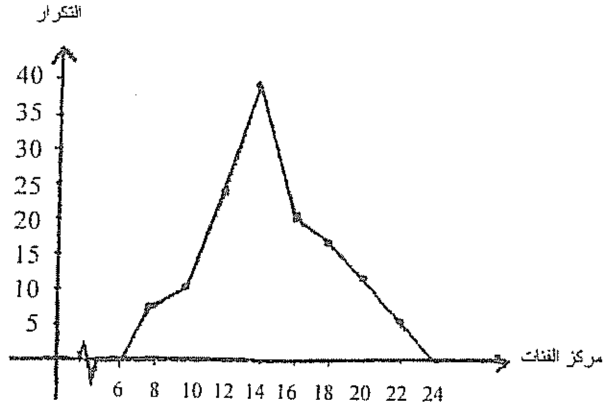


الشكل (1-20): توضيح للقاعدة التجريبية

مثال (18): بالعودة إلى المثال (10) حيث كان لدينا جدول التوزيع التكراري الفئوي التالي:

الفئة	21<23	19<21	17<19	15<17	13<15	11<13	9<11	7<9
التكرار	5	13	17	20	40	25	12	8

إنّ المضلع التكراري لهذا التوزيع يأخذ الشكل:



الشكل (1.21)

نلاحظ أنّ هذا الشكل قريب من شكل الجرس وبالتالي فإنّ توزيع القياسات في هذا المثال هو توزيع قريب من الطبيعي. لذلك يمكن تطبيق القاعدة التجريبية على هذا التوزيع.

لنحسب أولاً μ و σ حيث نعتبر أنّ تكرارات الفئات تمثل تكرارات مراكزها فيكون:

$$\mu = \frac{\sum cf}{\sum f} = \frac{2040}{140} = 14.6$$

$$\sigma^2 = \frac{\sum (c-\mu)^2 f}{\sum f - 1} = \frac{1647.4}{139} = 11.9$$

أي أنّ: $\sigma = 3.4$

وبالتالي وحسب القاعدة التجريبية نجد أنه يوجد تقريباً:

68.27٪ من القياسات في المجال $[\mu - \sigma \;,\; \mu + \sigma] = [11.2 \;,\; 18]$ وعددهم 96 قياس.

95.45٪ من القياسات في المجال $[\mu - 2\sigma \;,\; \mu + 2\sigma] = [7.8 \;,\; 21.4]$ وعددهم 134 قياس.

99.73٪ من القياسات في المجال $[\mu - 3\sigma \;,\; \mu + 3\sigma] = [4.4 \;,\; 24.8]$ وعددهم 140 قياس.

مثال (19): أخذنا عينة مؤلفة من 500 مصباح كهربائي من المصابيح المنتجة فــي إحــدى الشركات، حيث وُجد بعد فحصها أنها تعمل وسطياً لمدة 780 ساعة وبانحراف معياري قدره 38 ساعة، وبفرض أنّ شكل التوزيع هو شكل الجرس فالمطلوب استخدم القاعدة التجريبية في وصف هذه القياسات.

الحل: لدينا $\overline{x} = 780$ ، $s = 38$

فحسب القاعدة التجريبية يمكن أن نتوقع تقريباً أنّ:

68.27% من المصابيح المختبرة سوف تعمل مدة زمنية محصورة ضمن الفترة $[\overline{x}-s\ ,\ \overline{x}+s] = [742\ ,\ 818]$ وعددهم 341 مصباح.

95.45% من المصابيح المختبرة سوف تعمل مدة زمنية محصورة ضمن الفترة $[\overline{x}-2s\ ,\ \overline{x}+2s] = [704\ ,\ 856]$ وعددهم 477 مصباح.

99.73% من المصابيح المختبرة سوف تعمل مدة زمنية محصورة ضمن الفترة $[\overline{x}-3s\ ,\ \overline{x}+3s] = [666\ ,\ 894]$ وعددهم 499 مصباح.

1-8 تمارين عامة:

1- يبين الجدول التالي درجات 42 طالباً من طلاب السنة الثانية هندسة طاقـة فـي مقـرر الاحتمال والإحصاء في أحد الامتحانات:

36	48	45	65	45	27	40
50	35	52	58	56	40	63
52	60	65	25	55	51	45
37	36	50	65	40	48	62
55	69	25	45	26	58	35
45	74	27	35	37	52	55

والمطلوب:

1- أنشئ جدول التوزيع التكراري النقطي.

2- أنشئ جدول التوزيع التكراري للبيانات المبوبة مستخدماً 7 فئات متساوية الطول.

3- أنشئ المدرج التكراري والمضلع التكراري لهذا التوزيع.

4- أوجد التكرار النسبي لهذا التوزيع وأنشئ مضلعه التكراري.

5- أنشئ جدول التوزيع التكراري المجمع الصاعد وأرسم مضلعه التكراري.

6- أنشئ جدول التوزيع التكراري المجمع النازل وأرسم مضلعه التكراري.

7- أوجد المتوسط الحسابي النقطي والتبـاين والانحـراف المعيـاري النقطـي لهـذه الدرجات.

8- أوجد المتوسط الحسابي، والتباين والانحراف المعياري للبيانات المبوبة.

9- أوجد الوسط النقطي وكذلك المنوال النقطي لهذه القياسات.

10- أوجد الوسط والمنوال للبيانات المبوبة.

11- أوجد الربيعين الأدنى والأعلى للبيانات المبوبة ثم استنتج نصف المدى الربيعـي الفئوي لهذه البيانات.

2- قارن بين المجموعتين التاليتين من البيانات الممثلتين لعينتين مختلفتين مـن حيـث قـوة التمركز حول المتوسط الحسابي ودرجة التشتت.

A:	40	5	3	6	2
B:	31	29	32	30	33

3- لتكن البيانات التالية والتي تمثل عينة عشوائية مختارة:

40, 3, 6, 5, 7

والمطلوب حساب معاملي الالتواء الوسطي والتفلطح لهذه البيانات.

4- ليكن لدينا جدول التوزيع التكراري الفئوي التالي:

حدود الفئة	[20,23[	[23,26[	[26,29[	[29,32[	[32,35[	[35,38[	[38,41[
التكرار	4	8	23	34	26	4	1

والمطلوب:

1- أحسب المتوسط الحسابي، والتباين والانحراف المعياري للبيانات المبوبة.

2- أوجد الوسط وكذلك المنوال للبيانات المبوبة.

3- أوجد نصف المدى الربيعي لهذه البيانات.

4- أوجد المئوي P_{40}.

5- أحسب العُشير D_7.

5- إذا كان معدل الرواتب لعينة عشوائية مؤلفة من 1080 موظفاً هو 12000 ليرة سورية بانحراف معياري مقداره 800 ل.س. والمطلوب:

1- استخدم مبرهنة تشيبتشيف لتعيين المجال الذي يحوي على الأقل 810 موظفين من هذه العينة.

2- ما هو المجال الذي يمكن أن نكون متأكدين بأنه لا أكثر من 120 من الموظفين تقع رواتبهم خارجه.

6- أعلنت شركة لإنتاج أجهزة الحواسيب عن وجود 80 وظيفة شاغرة، فتقدم 1100 شخص بطلباتهم لشغل أياً من هذه الوظائف. وبعد إجراء الاختبار اللازم ولكل منهم، تبين أن متوسط مجموع درجات المتقدمين هو 175، بانحراف معياري قدره 10، فإذا علمت أنّ أحد المتقدمين حصل على مجموع قدره 215 فهل ستكون إحدى هذه الوظائف من نصيبه أم لا؟

7- تنتج إحدى شركات المصابيح الكهربائية، مصابيح تدعي أنها تعمل وسطياً لمدة 750 ساعة بانحراف معياري قدره 35 ساعة. ما هي النسبة المئوية العظمى لعدد المصابيح التي ستعمل لمدة أقل من 650 ساعة.

8- إذا علمت أنّ المعدل السنوي لسقوط الأمطار فوق إحدى المدن هو 52.75 cm بانحراف معياري قدره 3.45 cm أوجد النسبة المئوية العظمى للسنوات التي يتجاوز فيها المطـــر 70 cm.

9- تقدم 540 طالباً إلى الامتحان النهائي لمقرر الإحصاء والاحتمال. فكان المتوسط الحسابي لدرجاتهم 78 بانحراف معياري قدره 9 درجات. والمطلوب استخدم مبرهنة تشيبتشيف:

1- لتعيين المجال الذي يحوي على الأقل 405 طلاب من المتقدمين.

2- تعيين ذلك المجال الذي يمكن أن نحكم بأن لا أكثر من 60 طالباً يقع خارجه.

10- بفرض أنّ $x_1, x_2, \dots, x_n$ مجموعة من القياسات متوسطها μ وانحرافها المعيـــاري σ.
والمطلوب تحويل هذه القياسات إلى قياسات معيارية ثم إثبات أن:

1- متوسط القياسات المعيارية $z_1, z_2, \dots z_n$ هو الصفر.

2- تباين القياسات $z_1, z_2, \dots z_n$ يساوي الواحد.

11- أرادت إحدى شركات الغزل والنسيج أن توفد شخص للخارج لإتباع دورة تدريبية، فتقدم 6 مهندسين للاستفادة من هذه الدورة. أجرت الشركة ثلاثة اختبارات لكل من المتقدمين في مجال تخصصه كمهندس غزل وفي اللغة الإنكليزية وفي الحاسوب، فكانت النتـــائج كما يلي. (الدرجات من 10):

المهندس المتقدم	المعلومات ضمن تخصصه	اللغة الإنكليزية	الحاسوب
A	10	9	5
B	9	9	6
C	5	2	0
D	5	9	10
E	4	0	6
F	3	1	9

والمطلوب:

1- حول الدرجات إلى درجات معيارية في كلٍ من الاختبارات الثلاثة.

2- من هو الفائز بهذه الدورة، علماً بأنّ صاحب أكبر مجموع (بالواحدات المعيارية) هو الفائز.

12- تقدم طالب من طلاب السنة الثانية هندسة طاقة إلى ثلاثة امتحانات فكانت درجاته، بالإضافة إلى المتوسط الحسابي والانحراف المعياري لنتائج جميع الطلاب المتقدمين معه في كل امتحان من هذه الامتحانات الثلاث مسجلة في الجدول التالي:

المادة	الدرجة	المتوسط الحسابي	الانحراف المعياري
A	72	70.8	0.6
B	61	51.0	6.0
C	50	30.0	12.0

والمطلوب: في أي امتحان كانت نتيجة الطالب أفضل بالمقارنة مع زملائه.

13- تدفع شركة خاصة لموظفيها البالغ عددهم 1000 موظف رواتب شهرية بمعدل 14250 ل.س شهرياً بانحراف معياري قدره 600 ل.س. بفرض أن المنحني التكراري لمجتمع موظفي الشركة يأخذ شكل الجرس، استخدم القاعدة التجريبية لوصف فئات الموظفين تبعاً لتغير الرواتب.

14- لدى قياس أطوال 2000 طالب جامعي تبين أن المتوسط الحسابي لهذه الأطوال هو 174 cm. بانحراف معياري قدره 9cm. فإذا علمت أنّ منحني التوزيع التكراري لهذا المجتمع يأخذ شكل الجرس:

1- صف توزع هؤلاء الطلاب بالنسبة لأطوالهم مستخدماً القاعدة التجريبية.

2- كم طالباً يتجاوز طولهم 185 cm.

3- كم طالباً يقل طولهم عن 157 cm.

15- إذا كان المعدل الوسطي لحدوث أعطال كهربائية يومياً في منطقة ما هو 30 عطل بانحراف معياري قدره 5 أعطال، وبفرض أن المنحني التكراري له شكل الجرس. فالمطلوب:

1- صف توزع الأعطال مستخدماً القاعدة التجريبية.

2- أوجد النسبة المئوية لعدد الأيام التي لا يحدث فيها أكثر من 40 عطل.

3- أوجد النسبة المئوية لعدد الأيام التي يحدث فيها أقل من 17 عطل.

الفصل الثاني

مبادئ الاحتمال

2-1 – فضاء التجربة

2-2 – الحدث

2-3 – جبر الحوادث

2-4 – الفضاء الاحتمالي والدالة الاحتمالية

2-5 – خصائص الاحتمال

2-6 – الاحتمال الشرطي

2-7 – قانون بايز

2-8 – طرائق العد

2-9 – تمارين عامة

مبادئ الاحتمال
Elements of Probability

2-1 فضاء التجربة (Space of Experiment):

2-1-1 التجربة الاحصائية:

يستخدم في الإحصاء كلمة تجربة عشوائية للدلالة على أية عملية تقودنا إلى مجموعة من البيانات. فعند إلقاء قطعة نقدية في الهواء نجد أننا أمام نتيجتين منتظرتين «ظهور الصورة» أو «ظهور الكتابة» وعند سحب مصباح كهربائي من مجموعة مصابيح مخزنة نجد أن النتيجتين المنتظرتين «المصباح جيد» أو «المصباح تالف».

2-1-2 فضاء التجربة:

يطلق على مجموعة كل النتائج الممكنة لتجربة عشوائية أسم فضاء التجربة، ويُرمز لها بالرمز S. فمثلاً في عملية رمي قطعة نقود في الهواء نجد أن ناتج هذه التجربة هو إمّا ظهور الصورة H أو ظهور الكتابة T وعلى هذا يعبر عن فضاء هذه التجربة بالشكل:

$$S=\{H,T\}$$

وعند اختبار مصباح كهربائي، مأخوذ من مجموعة مصابيح فيما إذا كان يعمل أم لا، نجد أن فضاء التجربة في هذه الحالة هو:

$$S=\{D,N\}$$

حيث يدل الرمز D (Defective) على أن المصباح المسحوب تالف بينما يدل الرمز N (Non-defective) على أنه سليم.

في عملية إلقاء حجر النرد وبفرض أننا نهتم بالرقم الظاهر على الوجه العلوي للحجر نجد أن فضاء التجربة هو:

$$S=\{1,2,3,4,5,6\}$$

أمّا إذا كان اهتمامنا منصباً على صفة الرقم الظاهر «زوجي» أو «فردي» فإن فضاء التجربة عندئذٍ سيكون:

S={زوجي،فردي}

عند رمي ثلاث قطع نقود دفعة واحدة (أو رمي قطعة نقود ثلاث مرّات متتالية) نجد أن فضاء التجربة يتكون من كافة الثلاثيات التالية:

S={HHH,HTH,THH,HHT,TTH,THT,HTT,TTT}

وإذا كان إهتمامنا منصباً على عدد الصور الظاهرة خلال التجربة السابقة عندئذٍ يصبح فضاء التجربة كما يلي:

S={0,1,2,3}

أخيراً عند رمي حجري نرد في آن واحد فإننا نجد أن فضاء التجربة يتكون من كافة الثنائيات التالية:

S= {(1,1),(1,2),...,(6,6)}

والتي يكون عددها /36/ حيث يمكن استنتاج هذه الثنائيات بالاستفادة من الجدول التالي:

الجدول (2-1)

الحجر① \ الحجر②	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

2-1-3 الفضاء المنفصل والفضاء المتصل (Discrete and Continous Spaces):

تعريف (1):

نقول عن مجموعة A أنها مجموعة منتهية إذا كان عدد عناصرها منتهي، ونقول عن A أنها مجموعة قابلة للعد إذا كانت A مجموعة غير منتهية وبالإمكان عَدُّ عناصرها (أي يمكن مقابلة عناصرها مع عناصر مجموعة الأعداد الطبيعية). أمّا إذا كانت A مجموعة غير منتهية وغير قابلة للعد فإننا نقول عن A أنها مجموعة غير منتهية.

يُقال عن فضاء التجربة بأنه فضاءٌ منفصلٌ (منقطعٌ) إذا كان عدد عناصره منتهياً أو قابلاً للعد. كما يقال عن فضاء التجربة بأنه متصل (مستمر) إذا كان عدد عناصره غير منتهي (infinite) وغير قابل للعدّ (uncountable)، حيث يُقال في هذه الحالة عن التجربة الاحصائية بأنها تجربة احصائية متصلة وغير منفصلة. فمثلاً في تجربة قياس زمن احتراق مصابيح كهربائية نلاحظ أن هذه التجربة هي تجربة إحصائية متصلة وغير منفصلة لأن نواتج القياس قد تعطي أي عدد من الأعداد الحقيقية بين الصفر و 1000 ساعة مثلاً وهذه المجموعة من القياسات الاحصائية غير منتهية وغير قابلة للعدّ.

وبعبارة عامة يُقال عن فضاء التجربة بأنه فضاءٌ متصلٌ إذا كان بالإمكان الحصول على قياس بين كل قياسين أو على نتيجة إحصائية بين كل نتيجتين. مثل العمليات الاحصائية لقياس كميات الكهرباء المستهلكة، أو العمليات الاحصائية لقياس المسافة التي تتحركها سيارة بغالون واحد من البنزين.

تكتب عناصر فضاء التجربة المنفصل عادةً بالتفصيل حيث يتم حصرها بقوسين فمثلاً في تجربة اختبار آلة كهربائية فيما إذا كانت تعمل أم لا يكون فضاء التجربة بالشكل: $S=\{G,B\}$ حيث يدل الرمز G (Good) على أن الآلة تعمل، بينما يدل الرمز B (Bad) على أنها لا تعمل. بينما تكتب عناصر فضاء التجربة المتصل بصورة وصفية فمثلاً نلاحظ أن فضاء التجربة الناتج عن قياس زمن احتراق مصابيح كهربائية يكون بالشكل:

$$S=\{x : 0 \le x \le 1000 , x \in R\}$$

حيث تمثل R هنا الأعداد الحقيقية.

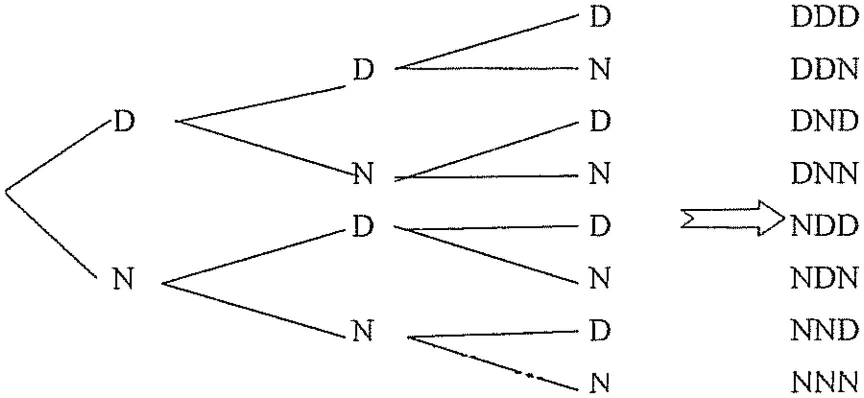
وفي كثير من الحالات نصادف تجارب عشوائية ذات صفة انفصالية (منقطعة) ولكن نتائجها كثيرة ولا نهائية بحيث يكون من الأفضل اعتبارها من التجارب المتصلة وذلك تسهيلاً للعمليات الحسابية والاستقراء الاحصائي. وهذا الاعتبار هو ما يسمى بعملية التقريب من الحالة المنقطعة إلى الحالة المتصلة (Approximation to continuity).

2-1-4 قواعد الاختيار والعدّ الاحصائي لكتابة فضاء التجربة:

في كثير من التجارب العشوائية الاحتمالية والتي تحتوي على عدد من المراحل تظهر الحاجة إلى حساب أو تعداد كافة نواتج التجربة أو بعبارة أخرى حساب فضاء التجربة بالكامل. حيث تستخدم طريقة الشجرة (Tree diagram) في مثل هذه الحالات خاصة عندما تتزايد مواصفات ومعالم الاختيار في التجربة العشوائية فمثلاً:

في عملية شراء ثلاث مصابيح من فئة الـ 100 واط لابد من اختبار المصابيح قبل الشراء حيث يتم هذا الاختبار وفقاً للطريقة أو الترتيب التالي:

ترتيب نواتج الاختبارات	الاختبار للمصباح الثالث	الاختبار للمصباح الثاني	الاختبار للمصباح الأول



حيث يدل الرمز D على أن المصباح تالف، بينما يدل الرمز N على أن المصباح سليم.

من طريقة الشجرة السابقة يمكننا معرفة التعداد الكلي لعناصر فضاء التجربة وهو:

$$N = |S| = n_1 \times n_2 .. \times n_k$$

حيث n_i يمثل عدد الحالات المتوقعة في الاختبار رقم i من جملة الاختبارات التي تتضمنها التجربة العشوائية.

فمثلاً في المثال السابق (اختبار ثلاثة مصابيح) نلاحظ أن عدد الحالات المتوقعة في كلٍ من الاختبارات الثلاثة هو (D أو N)

وبالتالي فإن عدد عناصر فضاء التجربة يكون:

$$N = |S| = 2 \times 2 \times 2 = 8$$

2-2 الحدث (Event):

عند إجراء تجربة عشوائية، يمكن أن يكون اهتمامنا موجهاً إلى نتيجة ما محددة من النتائج الممكنة للتجربة، أو إلى عدة نتائج محددة، أي نهتم بالحصول على عناصر محددة من فضاء التجربة. فمثلاً عند إلقاء حجر النرد يمكن أن نهتم بالحصول على الرقم 2 أو 4 أي نهتم فقط بالمجموعة {2,4} التي هي جزء من فضاء التجربة.

2-2-1 تعريف الحدث:

يُعرّف الحدث على أنه مجموعة جزئية من فضاء التجربة. كما يعرف الحدث الابتدائي(أو البسيط) على أنه مجموعة جزئية من فضاء التجربة ذات عنصر وحيد.

نستنتج من التعريف السابق أن فضاء التجربة S هو حدث لأن $S \subseteq S$، كذلك نجد أن المجموعة الخالية ϕ هي أيضاً حدث لأن $\phi \subset S$.

فمثلاً في تجربة إلقاء ثلاث قطع نقدية في الهواء نجد أن الحدث: «ظهور الكتابة على إحدى القطع والصورة على القطعتين الباقيتين» هو:

$$A = \{THH, HTH, HHT\}$$

والحدث "جميع الوجوه ذات شكل واحد" هو:

$$B = \{TTT, HHH\}$$

عند إجراء تجربة عشوائية ما مرةً واحدة فقط سنحصل على نتيجة واحدة فإذا فرضنا أن هذه النتيجة هي x فإننا نقول عن الحدث الاحصائي E بأنه قد وقع إذا وفقط إذا انتمت x إلى E. كما نقول عن أي حدث آخر تنتمي إليه x أنه قد وقع أيضاً.

نستنتج من ذلك أنه إذا كانت x نتيجة لتجربة عشوائية ما فإنه قد يقع أكثر من حدث في آن واحد، وأحد هذه الأحداث سيكون الحدث الابتدائي $\{x\}$.

فمثلاً عند رمي حجر النرد، إذا كان العدد الظاهر على الوجه العلوي هو 4 فإننا نقول بأن الحدث:

$$A = \{2,4,6\}$$

قد وقع. كذلك الأمر بالنسبة للحدث $B = \{3,4\}$ أو الحدث $C = \{4\}$ أو الحدث $S = \{1,2,3,4,5,6\}$.

لنفرض A,B حدثين موافقين لتجربة عشوائية ما، فإن $A \cap B$ و $A \cup B$ هي أحداث أيضاً حيث يقع الحدث $A \cap B$ إذا وفقط إذا وقع كل من A و B. بينما يقع الحدث $A \cup B$ إذا وفقط إذا وقع أيٌّ من A أو B (أي وقع أحدهما على الأقل).

فمثلاً عند سحب ورقة من أوراق الشدة (التي عددها 52 ورقة)، إذا كان A الحدث الذي يقع إذا كانت الورقة المسحوبة تحمل الرقم 5 وكان B الحدث الذي يقع إذا كانت الورقة المسحوبة من النوع الديناري فإن الحدث $A \cap B$ هو الحدث الذي يقع إذا كانت الورقة المسحوبة 5 ديناري بينما يكون $A \cup B$ الحدث الذي يقع إذا كانت الورقة المسحوبة تحمل الرقم 5 أو كانت دينارية.

2-2-2 الحدث الأكيد والحدث المستحيل:

بما أن الحدث هو مجموعة جزئية من فضاء التجربة وحيث أن $S \subseteq S$ فإن S هو حدث أيضاً وندعوه بالحدث الأكيد، لأن أي نتيجة نحصل عليها من تنفيذ التجربة العشوائية ستنتمي إلى S. كذلك نعلم بأن المجموعة الخالية $\phi \subset S$ أي أن ϕ حدث ولكن وقوعه مستحيل، لأن نتيجة التجربة لن تنتمي إلى ϕ. لذا ندعوه بالحدث المستحيل. فمثلاً في تجربة رمي حجر النرد نلاحظ أن فضاء التجربة $S = \{1,2,3,4,5,6\}$ هو حدث أكيد وأن ϕ حدث مستحيل.

2-2-3 الأحداث المتنافية (mutually Exclusive) والأحداث المتتامة:

تعريف (2): يُسمى الحدثان اللذان يستحيل وقوعهما معاً بالحدثين المتنافيين (وبمعنى آخر ينفي وقوع أحدهما وقوع الآخر) أي يكون الحدثان A و B متنافيين إذا كان: $A \cap B = \phi$

أمّا الحدثان المتتامان فهما الحدثان اللذان يقع أحدهما إذا وفقط إذا لم يقع الآخر. أي يكون الحدثان A و B متتامان إذا وفقط إذا كان:

$$A \cup B = S \,,\; A \cap B = \phi$$

أي $B = S - A = A'$

فمثلاً في تجربة رمي حجر النرد حيث نهتم بالرقم الظاهر على الوجه العلوي يكون فضاء التجربة S بالشكل التالي:

$$S = \{1,2,3,4,5,6\}$$

فإذا كانت A و B و C الأحداث الإحصائية التالية:

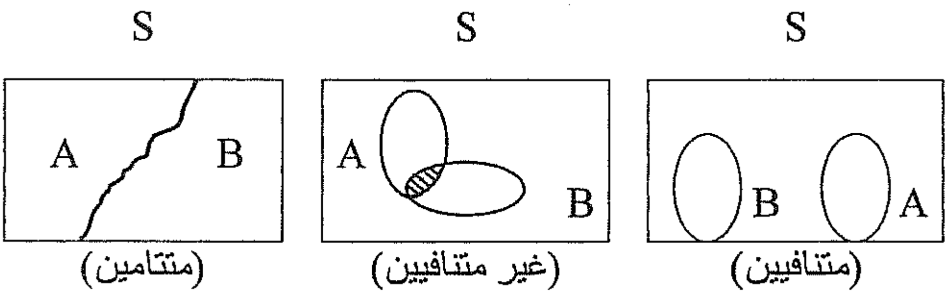
$$A = \{1,3,5\}$$
$$B = \{2,4,6\}$$
$$C = \{2,4\}$$

فعندئذٍ نلاحظ أن الحدثين A و C متنافيان لأن: $A \cap C = \phi$ كما نلاحظ أن الحدثين A,B متتامان لأن

$$A \cup B = S \,,\; A \cap B = \phi$$

ممّا سبق نستنتج أن الحدثين المتتامين هما حدثان متنافيان والعكس غير صحيح بالضرورة.

ملاحظة ①: يمكن توضيح مفهوم فضاء التجربة والحوادث بمخططات فن (Venn) حيث يتم تمثيل فضاء التجربة S بمستطيل وتمثيل الحوادث بدوائر أو أشكال بيضوية أو مناطق جزئية واقعة داخل هذا المستطيل. الشكل التالي يبين بعض الحالات الممثلة لنوع الارتباط بين حدثين مفروضين A و B:



الشكل (2-1): توضيح العلاقة بين حدثين احصائيين مفروضين

2-2-4 الأحداث المستقلة (Independent events):

تسمى الأحداث التي يكون حصولها أو وقوعها غير مترابط ببعضها البعض ضمن تجربة عشوائية ما باسم الأحداث المستقلة وبعكس ذلك أي الأحداث التي يترابط وقوعها ببعضها البعض اسم الأحداث غير المستقلة أو الأحداث المترابطة (dependent events).

فمثلاً نجد أن هطول المطر مرتبط بالغيوم وحركة الغيوم ترتبط بالرياح لذا فهذه الأحداث الجوية الثلاثة غير مستقلة. وفي تجربة رمي قطعة نقود متزنة نجد أن ظهور الصورة ينفي ظهور الكتابة وبالتالي نلاحظ أن ظهور الكتابة متعلق بعدم ظهور الصورة لذا فالحدثان (ظهور الصورة وظهور الكتابة) غير مستقلين على الرغم من كونهما متنافيين.

أمّا ظهور الصورة H على قطعة نقود متزنة لا ينفي ولا يتعلق بظهور صورة أخرى H على قطعة نقود أخرى وعلى هذا الأساس نجد أن الحدثين ظهور H على قطعة نقود ما، وظهور H على قطعة أخرى هما حدثان مستقلان. وبالمثل ظهور العدد 5 عند رمي حجر نرد وظهور الصورة H عند رمي قطعة نقود، هما حدثان مستقلاّن.

وكذلك النجاح في مادة الإحصاء والنجاح في مادة الجغرافيا هما حدثان مستقلان. واصطدام سيارة في شارع ما من مدينة دمشق وأخرى في شارع ما من مدينة حمص، هما حدثان مستقلان. وهكذا...

نتيجة (1): مما سبق نجد أن:

أ – الحدثان المتنافيان هما حدثان غير مستقلين.

ب – الحدثان المستقلان هما حدثان غير متنافيين.

جـ – أمّا الأحداث غير المتنافية فقد تكون مستقلة أو غير مستقلة. وكذلك الأحداث غير المستقلة فقد تكون متنافية أو غير متنافية.

فمثلاً ظهور الصورة H في الرمية الأولى لقطعة نقود وظهورها (أي ظهور H) في الرمية الثانية يعبّر عن حدثين غير متنافيين ولكنهما مستقلان.
كذلك نجد أن ظهور الملكة عند سحب ورقة لعب من أوراق الشدة وعددها 52، لا ينفي ظهور الملك في السحبة الثانية إذا فرضنا أنّ السحب يتم مع عدم إرجاع الكرت المسحوب. حيث نجد أن هذين الحدثين المذكورين غير متنافيين ولكنهما غير مستقلين أيضاً.

2-3 جبر الحوادث (Algebra of Events):

لنوضح أولاً وبشكل موجز ماذا نعني بجبر المجموعات بشكل عام؟. لتكن S مجموعة مفروضة عندئذٍ نسمي أي مجموعة عناصرها مجموعات جزئية من S بالصف، حيث تكون مجموعة أجزاء S وهي $P(s)$ أحد هذه الصفوف. فإذا كان لدينا على سبيل المثال المجموعة:

$$S = \{1,2,3\}$$

فإنه يمكننا تشكيل العديد من الصفوف من هذه المجموعة منها:

$$\zeta_1 = \{\{1\},\{2,3\}\} \quad , \quad \zeta_2 = \{\{1\},\{2\}\} \quad , \quad \zeta_3 = \{\phi, \{1\},\{2,3\},S\}\ldots$$

حيث نلاحظ من الصفوف السابقة أن الصفين ζ_1 و ζ_2 غير مغلقين بالنسبة للعمليات الجبرية المنطقية وذلك لأن:

$\{1\}\in\zeta_1 , \{2,3\}\in\zeta_1$ ولكن $\{1\}\cup\{2,3\}=\{1,2,3\}\notin\zeta_1$ كما أن $\{1\}\in\zeta_2$ ولكن $(\{1\})' = \{2,3\}\notin\zeta_2$ بالاضافة لكون $S\notin\zeta_1$ و $S\notin\zeta_2$. بينما نلاحظ أن الصف ζ_3 هو صف من أجزاء S لا يساوي $P(S)$ ولكنه يحتوي على المجموعة S كما أنه مغلق بالنسبة للعمليات المنطقية، لأن اجتماع أو تقاطع أي جماعة من عناصره هو عنصر ينتمي إليه، كما أن متممة أي عنصر منه تنتمي إليه.
إن من المهم في حساب الاحتمالات وغيره من الفروع الرياضية أن تكون صفوف المجموعات التي نعمل عليها مغلقة بالنسة للعمليات المنطقية سواءً كانت هذه العمليات مطبّقة عدداً منتهياً من المرات أو معدوداً في حالات أخرى.

تعريف (3): إذا كانت S مجموعة مفروضة وكان ξ صفاً غير خالٍ من أجزاء S$(\xi \subseteq P(S))$ فإننا نقول عن ξ أنه جبراً على S إذا تحققت الشروط التالية:

1- $S\in\xi$

$$A \in \xi \Rightarrow A' \in \xi \qquad -2$$

$$A, B \in \xi \Rightarrow A \bigcup B \in \xi \qquad -3$$

فإذا كان جبراً وحقق إضافة إلى ذلك الشرط التالي:

$$A_1, A_2, \ldots, A_n, \ldots \in \xi \Rightarrow \bigcup_{i \geq 1} A_i \in \xi \qquad -4$$

الذي نعبر عنه بقولنا أن ξ مغلق بالنسبة للاجتماع المعدود، فإننا نقول عن ξ أنه جبر تام على S.

مثال (1): إن $P(S)$ جبر وجبر تام.

مثال (2): إن $\xi = \{\phi, S\}$ جبر كما أنه جبر تام.

مثال (3): إذا كانت $\xi = \{\phi, \{1\}, \{4\}, \{2,3\}, \{1,4\}, \{1,2,3\}, \{2,3,4\}, S\}$
حيث $S = \{1,2,3,4\}$ ، فمن السهل التحقق من أن هذا الصف جبر وجبر تام.

مثـال (4): إذا كانت $S = \{1,2,3,4\}$ فمن الواضح أن الصـف $\zeta = \{\{1\}, \{2,3\}\}$ لـيس جبراً.

تعريف (4): إذا كانت S تمثل فضاء تجربة ما وكان ξ صفاً غير خال من أجزاء S بحيث يحقق شروط الجبر فإننا نسمي ξ جبراً من الحوادث على S.

2-4 الفضاء الاحتمالي والدالة الاحتمالية :(Probability Space and Probability Function)

يعرف الاحتمال عادةً بأنه درجة الظن في وقوع حدث ما. فإذا كان S فضاء تجربة ما يحتوي على n نتيجة كل منها لها نفس الفرصة في الظهور، وكان A حدثاً يحتوي على k نتيجة فإن احتمال وقوع الحدث A (أو بتعبير آخر تحققه) هو:

$$P = P(A) = \frac{|A|}{|S|} = \frac{k}{n} = \frac{\text{عدد الحالات المواتية (الموافقة لـ } A)}{\text{عدد الحالات الكلية}}$$

وينتج عن ذلك أن احتمال عدم وقوع الحدث A (أو بتعبير آخر عدم تحققه) هو احتمال وقوع الحدث المتمم لـ A. أي أن:

$$q = P(A') = \frac{n-k}{n} = 1 - \frac{k}{n}$$

أي بتعبير آخر: $P(A') = 1 - p(A)$ وبهذا فإن $P(A) + P(A') = 1$ أو $P+q = 1$

مثال (5): عند سحب ورقة لعب من أوراق الشدة التي عددها 52 فإن احتمال وقوع الحدث A الذي يقع إذا و فقط إذا "كانت الورقة المسحوبة دينارية" هو:

$$p(A) = \frac{13}{52} = \frac{1}{4}$$

كما أن احتمال وقوع الحدث B الذي يقع إذا وفقط إذا "كانت الورقة المسحوبة تحمل الرقم 6" هو:

$$p(B) = \frac{4}{52} = \frac{1}{13}$$

أمّا احتمال أن لا تكون الورقة المسحوبة تحمل الرقم 6 فهو:

$$P(B') = \frac{48}{52} = \frac{12}{13}$$

إن التعريف التقليدي السابق للاحتمال غير كافٍ في حساب الاحتمالات فهو يشترط أن تكون كل نتيجة من نتائج فضاء التجربة لها نفس الفرصة في الظهور، لذلك نحتاج إلى تعريف أعم وأدق من التعريف السابق بحيث يمكن من خلاله إيجاد احتمال حدث ما حتى عندما تكون نتائج فضاء التجربة ليس لها نفس الفرص في الظهور.

التعريف الرياضي للاحتمال:

تعريف (4): الفضاء الاحتمالي ثلاثية (S, ξ, P) فيها S مجموعة غير خالية و ξ جبر حوادث تام من أجزاء S و P دالة لـ ξ في R (مجموعة الأعداد الحقيقية) تحقق الخصائص التالية:

1- $0 \le P(A) \le 1$ وذلك مهما يكن A من ξ.

2- $P(S) = 1$

3- إذا كانت $A_1, A_2, \ldots, A_n, \ldots$ متتالية منتهية أو غير منتهية من عناصر ξ المتنافية مثنى مثنى كان:

$$P\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i\geq 1} P(A_i)$$

وعندئذ نقول عن الدالة P أنها دالة احتمالية وعن $P(A)$ ، قيمة الدالة P عند A ، أنها احتمال وقوع الحدث A.

نتيجة (2):

1- اعتماداً على التعريف السابق للدالة الاحتمالية نستنتج أنه إذا كانت $A_1,A_2...,A_n$ مجموعة منتهية من الأحداث المتنافية مثنى مثنى فإن:

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n) \qquad \textbf{(2.1)}$$

2- إذا كان فضاء التجربة S يحتوي على n نتيجة $x_1, x_2, ..., x_n$ وبما أن الحوادث الابتدائية $\{x_1\},\{x_2\}...\{x_n\}$ تشكل جماعة من الحوادث المتنافية مثنى فإن:

$$P(S) = P(\bigcup_{i=1}^{n}\{x_i\} = \sum_{i=1}^{n} P(\{x_i\})$$

فإذا رمزنا للاحتمال $P(\{x_i\})$ بالرمز $P(x_i)$ اختصاراً، من أجل الاحتمال لحدث ابتدائي وبملاحظة أن $P(S)=1$ نجد أن: $1 = \sum_{i=1}^{n} P(x_i)$

حيث نستنتج في الحالة التي تكون فيها جميع الحوادث الابتدائية $\{x_i\}$ لها نفس الفرصة في الظهور (متساوية الاحتمال)، أن:

$$1 = nP(x_i) \Rightarrow P(x_i) = \frac{1}{n}$$

وبالتالي فإن احتمال أي حدث A عدد عناصره k في هذه الحالة هو $P(A) = \frac{k}{n}$

وهذا ما وجدناه من خلال التعريف التقليدي للاحتمال مع التذكير بأن ما ذكر أعلاه يبقى مشروطاً بالحالة التي تكون فيها جميع نتائج فضاء التجربة متساوية الفرصة في الظهور.

مثال (6): عائلة عندها ثلاثة أطفال، ما احتمال أن يكون عندها صبيّان وبنت علماً بأن للبنت وللصبي نفس الفرصة بالولادة؟

الحل: إن فضاء التجربة هو:

$$S = \{GGG, GGB, GBG, BGG, GBB, BGB, BBG, BBB\}$$

وعناصر S كلها متساوية الاحتمال. إن الحدث المطلوب هو:

$$A = \{BBG, BGB, GBB\}$$

وبالتالي فإن:

$$P(A) = \frac{3}{8}$$

مثال (7): عند إلقاء قطعة نقود متوازنة مرتين متتاليتين. أوجد احتمال ظهور صورة واحدة على الأقل.

الحل: بما أن قطعة النقود متوازنة فإن «ظهور الكتابة» و «ظهور الصورة» متساوياً الفرص في الظهور. إن فضاء التجربة هو:

$$S = \{HH, HT, TH, TT\}$$

حيث تكون جميع نتائج S متساوية الاحتمال. إن الحدث المطلوب في هذا المثال هو:

$$A = \{HT, TH, HH\}$$

وبالتالي فإن :

$$P(A) = \frac{3}{4}$$

مثال (8): أراد شخص استخدام عدد من أربع خانات لتشكيل عددٍ سريٍّ للدخول إلى حاسبه الشخصي ما احتمال أن يحتوي هذا العدد على ثلاثة أرقام زوجية ورقم فردي علماً بأن للأرقام نفس الفرصة في الاختيار.

الحل: إذا رمزنا بالرمز E للرقم الزوجي والرمز O للرقم الفردي فإن فضاء التجربة يكون:

$$S = \{EEEE, EEEO, EEOE, EEOO, EOEE, EOEO, EOOE, EOOO, OEEE, OEEO, OEOE, OEOO, OOEE, OOEO, OOOE, OOOO\}$$

حيث يكون الحدث المطلوب هو:

$$A = \{EEEO, EEOE, EOEE, OEEE\}$$

وبما أن للأرقام نفس الفرصة في الاختيار فإن:

$$P(A) = \frac{4}{16} = \frac{1}{4}$$

مثال (9): ألقي حجر نرد بحيث أن فرصة ظهور العدد الزوجي تساوي ضعف فرصة ظهور العدد الفردي. ما احتمال الحدث A الذي يقع إذا وفقط إذا كان الناتج أكبر من 3. وما هو احتمال الحدث B الذي يقع إذا وفقط إذا كان الناتج يقبل القسمة على 3. ثم ما هو احتمال الحدث C الذي يقع إذا وفقط إذا كان الناتج عدداً زوجياً؟

الحل: إن فضاء التجربة S هو:

$$S = \{1,2,3,4,5,6\}$$

فإذا فرضنا أن احتمال ظهور العدد الفردي هو x فإن احتمال ظهور العدد الزوجي يكون 2x وبالتالي فإن:

$$P(S) = P(1) + P(2) + P(3) + P(4) + P(5) + P(6)$$
$$1 = x + 2x + x + 2x + x + 2x$$
$$1 = 9x \rightarrow x = \frac{1}{9}$$

وبالعودة لنص المثال نلاحظ أنه لدينا الأحداث التالية:

$$A = \{4,5,6\}$$
$$B = \{3,6\}$$
$$C = \{2,4,6\}$$

حيث نجد أن:

$$P(A) = P(4) + P(5) + P(6) = \frac{2}{9} + \frac{1}{9} + \frac{2}{9} = \frac{5}{9}$$
$$P(B) = P(3) + P(6) = \frac{1}{9} + \frac{2}{9} = \frac{3}{9} = \frac{1}{3}$$
$$P(C) = P(2) + P(4) + P(6) = \frac{2}{9} + \frac{2}{9} + \frac{2}{9} = \frac{6}{9} = \frac{2}{3}$$

2-5 خصائص الاحتمال:

من التعريف الرياضي للدالة الاحتمالية يمكننا ايجاد الخصائص التالية:

أ – خاصة ضرب الاحتمالات للأحداث المستقلة والغير متنافية:

احتمال تقاطع حدثين مستقلين وغير متنافيين يقعان معاً يساوي حاصل ضرب احتمالي وقوع كل منهما. أي أن:

$$P(A \cap B) = P(A).P(B) \qquad (2.2)$$

وبالمثل يمكن توسيع القاعدة لتشمل n من الأحداث المستقلة والغير متنافية حيث نجد أن:

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)...P(A_n) \qquad (2.3)$$

فمثلاً في تجربة رمي قطعة نقود متوازنة ثلاث مرّات نجد أن ظهور الصورة H في الرمية الأولى وظهورها في الرمية الثانية وظهورها في الرمية الثالثة أحداث مستقلة وغير متنافية وإذا عبّرنا عن وقوعها معاً بالرمز HHH فإن:

$$P(HHH) = P(H)\ P(H)\ P(H) = \frac{1}{2}.\frac{1}{2}.\frac{1}{2} = \frac{1}{8}$$

بنفس الطريقة نجد أنّ:

$$P(HTH) = \frac{1}{2}.\frac{1}{2}.\frac{1}{2} = \frac{1}{8}$$

ب – إذا كان A و B حدثان فإن:

$$P(B) = P(B \cap A) + P(B \cap A') \qquad (2.4)$$

الاثبات: ليكن A حدث ما و A' الحدث المتمم له عندئذ نعلم بأن:

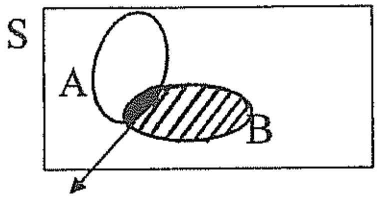
$$S = A \cup A'$$

وبفرض B حدث آخر عندئذ يكون:

$$B = B \cap S = B \cap (A \cup A') = (B \cap A) \cup (B \cap A')$$

وبما أن الحدثان $B \cap A$ و $B \cap A'$ متنافيان ومن العلاقة (2.1) نجد أن:

$$P(B) = P(B \cap A) + P(B \cap A')$$



الشكل (2-2)

جـ -

$$1 = P(A) + P(A') \qquad \textbf{(2.5)}$$

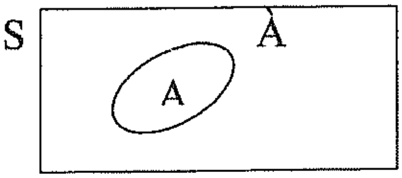
الاثبات:

ليكن الحدث A وليكن A' الحدث المتمم له عندئذ بما أن: $S = A \cup A'$ وبملاحظة أن A و A' حدثان متنافيان والاستفادة من العلاقة (2.1) نجد أن:

$$P(S) = P(A \cup A') = P(A) + P(A')$$

ولكن $P(S) = 1$ من تعريف الدالة الاحتمالية وبالتالي نجد أن:

$$1 = P(A) + P(A')$$



الشكل (2-3)

يمكن اثبات العلاقة (2.5) أيضاً بطريقة ثانية وذلك باستفادة من العلاقة (2.4)، فإذا وضعنا $S = B$ في تلك العلاقة نجد أن:

$$P(S) = P(S \cap A) + P(S \cap A')$$

ومنه:

$$1 = P(A) + P(A')$$

نستنتج من العلاقة (2.5) أنّ:

$$P(A') = 1 - P(A)$$

وكذلك:

$$P(A) = 1 - P(A')$$

ء –

$$P(\phi) = 0 \qquad \textbf{(2.6)}$$

الاثبات:

إذا وضعنا $A = \phi$ في العلاقة (2.5) نجد أن:

$$1 = P(\phi) + P(\phi')$$

ومنه:

$$1 = P(\phi) + P(S)$$

وبما أن $P(S) = 1$ فإن: $P(\phi) = 0$

هـ- إذا كان A و B حدثان فإن:

$$P(A \cap B') = P(A) - P(A \cap B) \qquad \textbf{(2.7)}$$

الاثبات:

من الشكل (4-2) نلاحظ أنّ:

$$A = (A \cap B') \cup (A \cap B)$$

وبملاحظة أن الحدثين $A \cap B$ و $A \cap B'$ هما حدثان متنافيان لذلك وبالاستفادة من العلاقة (2.1) نجد أنّ:

$$P(A) = P\ [(A \cap B') \cup (A \cap B)]$$

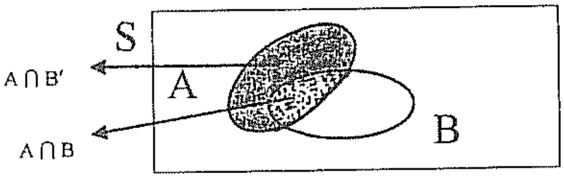
$$P(A) = P\ (A \cap B') + P(A \cap B)$$

أي أن:

$$P(A \cap B') = P(A) - P(A \cap B)$$

نستنتج كذلك من العلاقة (2.7) أنّ:

$$P(A) = P(A \cap B) + P(A \cap B') \qquad (2.8)$$



الشكل (2-4)

و – إذا كان A و B حدثان غير متنافيين فإن:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (2.9)$$

الاثبات:

من الشكل (2-5) نلاحظ أن الحدثين B و $A \cap B'$ هما حدثان متنافيان لأنّ:

$$(A \cap B') \cap B = \phi$$

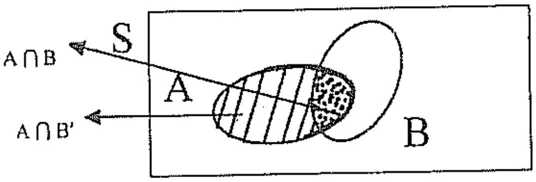
وبما أنّ:

$$(A \cup B) = (A \cap B') \cup B$$

فإنّ:

$$P(A \cup B) = P(A \cap B') + P(B)$$

وذلك بالاستفادة من العلاقة (2.1)



الشكل (2-5)

وبملاحظة العلاقة (2.7) نجد أنّ:

$$P(A \cup B) = P(A) - P(A \cap B) + P(B)$$

أي أنّ:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

ز – إذا كان A و B حدثان وكان: $A \subset B$ فإنّ:

$$P(A) \leq P(B)$$

الاثبات:

من الشكل (6-2) نلاحظ أنّ:

$$B = A \cup (A' \cap B)$$

وبما أن الحدثين: A و $A' \cap B$ متنافيان لأنّ:

$$A \cap (A' \cap B) = \phi$$

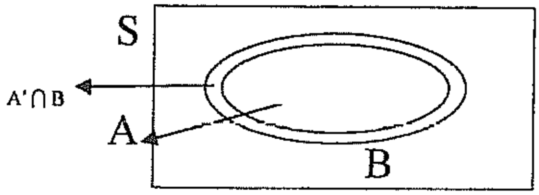
فإنّ:

$$P(B) = P\ [A \cup (A' \cap B)]$$
$$= P(A) + P(A' \cap B)$$

وذلك بالاستفادة من العلاقة (2.1). ومنه: $P(B) - P(A) = P(A' \cap B) \geq 0$

أي أنّ:

$$P(B) \geq P(A)$$



الشكل (6-2)

مثال (10): أكتب فضاء التجربة لأربعة متسابقين إذا علمت أن نتيجة المسابقة هي النجاح أو الفشل، ثم أوجد احتمال أن يكون المتسابق الثاني ناجحاً، وكذلك احتمال أن يكـون المتسـابق الثالث فاشلاً وأخيراً أوجد احتمال أن يكون المتسابق الثاني ناجحاً والثالث فاشلاً.

الحل: إذا رمزنا بالرمز p (passe) للنجاح والرمز f (failier) للفشل فإن فضـاء التجربـة يكون:

S = {pppp, pppf, ppfp, pfpp, fppp, ppff, pfpf, pffp, fpfp, ffpp, fppf, pfff, fpff, ffpf, fffp, ffff}

بفرض أنّ A هو الحدث الذي يقع إذا وفقط إذا كان المتسابق الثاني ناجحاً و B هو الحدث الذي يقع إذا كان المتسابق الثالث فاشلاً فإن احتمال أن يكون المتسابق الثاني ناجحاً هو :

$$P(A)=\frac{8}{16}=\frac{1}{2}$$

واحتمال أن يكون المتسابق الثالث فاشلاً هو:

$$P(B)=\frac{8}{16}=\frac{1}{2}$$

وبما أنّ الحدثين A و B مستقلان وغير متنافيين فإنّ:

$$P(A\cap B)=P(A).P(B)=\frac{1}{2}.\frac{1}{2}=\frac{1}{4}$$

مثال (11): سحبنا ورقة لعب من شدة (عدد أوراقها 52). ما هو احتمال أن تكون دينارية أو ملك.

الحل: ليكن A الخدث الذي يقع إذا وفقط إذا كانت الورقة المسحوبة دينارية، و B الحدث الذي يقع إذا وفقط إذا كانت الورقة المسحوبة ملك. وبالتالي فإن $A\cap B$ هو الحدث الذي يقع إذا وفقط إذا كانت الورقة المسحوبة ملك ديناري. حيث نلاحظ أنّ:

$$P(A)=\frac{13}{52}=\frac{1}{4}$$

$$P(B)=\frac{4}{52}=\frac{1}{13}$$

$$P(A\cap B)=\frac{1}{52}$$

وبما أنّ الحدثين A و B غير متنافيين فإنّ:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{1}{4}+\frac{1}{13}-\frac{1}{52}=\frac{4}{13}$$

مثال (12): إذا كان احتمال نجاح أحمد هو $\frac{1}{4}$ واحتمال رسوب باسل هو $\frac{1}{5}$ واحتمال نجاح الاثنين معاً هو $\frac{1}{8}$، فما هو احتمال نجاح أحدهما على الأقل؟

الحل: نفرض A الحدث الذي يقع إذا نجح أحمد و B الحدث الذي يقع إذا نجح باسل وبالتالي فإن الحدث المطلوب حساب احتماله هو $A\cup B$. حيث لدينا:

$$P(A) = \frac{1}{4}$$

$$P(B) = 1 - P(B') = 1 - \frac{1}{5} = \frac{4}{5}$$

وبما أن: $P(A \cap B) = \frac{1}{8}$

فإن الحدثين B و A غير متنافيين وبالتالي فإن:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{1}{4} + \frac{4}{5} - \frac{1}{8} = \frac{37}{40}$$

مثال (13): إذا كان احتمال إصابة محمد لهدف معين هو $\frac{3}{4}$ واحتمال إصابة فايز لنفس الهدف هو $\frac{1}{3}$، فاحسب كلاً من الاحتمالات التالية:

أن لا يصيب محمد الهدف، أن يصيبا الهدف معاً، أن يصيب أحدهما على الأقل، أن لا يصيبا الهدف في آنٍ واحد (أي أن يصيب أحدهما فقط أو لا يصيبا معاً)، أن يصيب محمد الهدف ولا يصيبه فايز.

الحل: نفرض أنّ A الحدث الذي يقع إذا أصاب محمد الهدف و B الحدث الذي يقع إذا أصاب فايز الهدف عندئذٍ نجد أنّ:

$$P(A) = \frac{3}{4} \quad و \quad P(B) = \frac{1}{3}$$

وبالتالي فإنّ:

احتمال أن لا يصيب محمد الهدف هو: $P(A') = 1 - P(A) = 1 - \frac{3}{4} = \frac{1}{4}$

واحتمال أن يصيبا الهدف معاً هو: $P(A \cap B) = P(A)P(B) = \frac{3}{4} . \frac{1}{3} = \frac{1}{4}$

وذلك بملاحظة أن B و A حدثين مستقلين وغير متنافيين.

أمّا احتمال أن يصيب الهدف أحدهما على الأقل فهو:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{4} + \frac{1}{3} - \frac{1}{4} = \frac{5}{6}$$

ولحساب احتمال أن لا يصيبا الهدف في آنٍ واحد يجب أن نحسب الاحتمال للحدث $(A \cap B)'$ حيث نجد أنّ:

$$P(A\cap B)'=1-P(A\cap B)=1-\frac{1}{4}=\frac{3}{4}$$

وأخيراً نجد أن احتمال أن يصيب محمد الهدف ولا يصيبه فايز هو:

$$P(A\cap B')=P(A)-P(A\cap B)=\frac{3}{4}-\frac{1}{4}=\frac{1}{2}$$

2-6 الاحتمال الشرطي (Conditional Probability):

من المسائل الهامة في حساب الاحتمالات دراسة العلاقة ما بين الأحداث المتعلقة بتجربة مفروضة. إن للأحداث المتعلقة بتجربة مفروضة احتمالاتها قبل إجراء التجربة، ولكن إذا أُجريت هذه التجربة وتلقّينا إعلاماً جزئياً عن النتيجة التي وقعت كأن نعلم أنّ حدثاً معيناً قد وقع، فإن من شأن ذلك أن تتغير الصورة الاحتمالية وأن تتغير بالتالي أحكامنا الاحتمالية الأخرى، فيزيد احتمال بعض الأحداث وينقص احتمال بعضها الآخر. وقد توجد أحداث لا تتغير احتمالاتها فتبقى كما كانت عليه قبل إجراء التجربة.

وبشكل عام إذا كان A و B حدثان متعلقان بتجربة مفروضة، فإننا نعني باحتمال A مشروطاً بالحدث B والذي نرمز له بالرمز $P(A|B)$، احتمال حدوث A علماً بأن B قد وقع. أي أن $P(A|B)$ يمثّل إعادة تقييمنا لاحتمال A على ضوء الاعلام الذي نتلقاه حول وقوع B.

هذا ويمكن توضيح مفهوم الاحتمال الشرطي من خلال المثال التالي:

مثال (14): في تجربة إلقاء حجر نرد متوازن، ليكن A الحدث الذي يقع إذا وفقط إذا كان الناتج فردياً. والمطلوب:

1- ما هو احتمال وقوع الحدث A.

2- ألقي حجر النرد فكان الناتج أكبر من 3. ما احتمال وقوع A في هذه الحالة.

الحل: نفرض B الحدث الذي يقع إذا وفقط إذا كان الناتج أكبر من 3. عندئذٍ نجد أنّ:

1- احتمال الحدث A قبل إلقاء حجر النرد ووقوع الحدث B هو:

$$P(A)=\frac{3}{6}=\frac{1}{2}$$

مع التذكير بأن فضاء التجربة هنا هو: $S=\{1,2,3,4,5,6\}$

2- بعد إلقاء حجر النرد ووقوع الحدث B تصبح مجموعة النتائج الممكنة هي: $B = \{4,5,6\}$ وهذا يعني أنّ B هو فضاء التجربة الجديد وبالتالي فإنّ احتمال الحدث A بعد وقوع B هو:

$$P(A|B)=\frac{1}{3}$$

تعريف (5): إذا كان (S, ξ, p) فضاءً احتمالياً وكان B حدثاً احتماله لا يساوي الصفر و A حدثاً اختيارياً فإنّ الاحتمال الشرطي للحدث A علماً بأن B قد وقع هو بالتعريف:

$$P(A|B)=\frac{P(A\cap B)}{P(B)} \qquad \textbf{(2.10)}$$

مثال (15): الجدول التالي يمثل توزيع العاملين في مصنع ما تبعاً للجنس والحالة الوظيفية.

الجنس/العمل	عامل	إداري	المجموع
ذكر	78	22	100
أنثى	46	34	80
المجموع	124	56	180

أختير شخص عشوائياً من هذا المصنع. ما احتمال أن يكون رجلاً علماً بأنه عامل، وما احتمال أن يكون الشخص أنثى من الإداريين؟

الحل: نفرض A الحدث الذي يقع إذا وفقط إذا كان الشخص رجلاً.

B الحدث الذي يقع إذا وفقط إذا كان الشخص عاملاً.

عندئذٍ نلاحظ أنّ: $P(B)=\frac{124}{180}=\frac{31}{45}$

وأنّ: $P(A\cap B)=\frac{78}{180}=\frac{39}{90}$

وبالتالي فإنّ الاحتمالين المطلوبين هما:

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{39}{90}}{\frac{31}{45}}=\frac{39}{90}.\frac{45}{31}=\frac{39}{62}=0.629$$

$$P(A'|B')=\frac{P(A'\cap B')}{P(B')}=\frac{\frac{34}{180}}{\frac{56}{180}}=\frac{34}{56}=\frac{17}{28}=0.607$$

1-6-2- خصائص الاحتمال الشرطي:

من تعريف الاحتمال الشرطي المبيّن بالعلاقة (2.10) يمكننا استنتاج الخواص التالية:

أ) إذا كان A و B حدثان من الفضاء الاحتمالي (S, ξ, P) و $P(A) > 0$ فإنّ:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad \textbf{(2.11)}$$

الاثبات:

من تعريف الاحتمال الشرطي مباشرةً نجد أنّ:

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$= \frac{P(A \cap B)}{P(A)}$$

وذلك لأن التقاطع عملية تبديلية.

ب) إذا كان A و B حدثان من الفضاء الاحتمالي (S, ξ, P) فإنّ احتمال تقاطعهما:

$$P(A \cap B) = P(A).P(B|A) = P(B).P(A|B) \qquad \textbf{(2.12)}$$

حيث يمكن استنتاج ذلك مباشرةً من العلاقتين (2.10) و (2.11). وبالمثل إذا كانــت A و B و C ثلاث حوادث من الفضاء الاحتمالي (S, ξ, P) فإنّ:

$$P(A \cap B \cap C) = P(A).P(B|A)P(C|(A \cap B)) \qquad \textbf{(2.13)}$$

وبهذه الطريقة يمكننا حساب احتمال التقاطع لأي عدد منتهي من الحــوادث: $A_1, A_2, \ldots, A_n$ بالشكل:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = \qquad \textbf{(2.14)}$$
$$P(A_1)P(A_2|A_1)P(A_3|(A_1 \cap A_2))\ldots P(A_n|(A_1 \cap A_2 \ldots \cap A_{n-1}))$$

جـ) – إذا كان A و B حدثان متنافيان فإنّ:

$$P(A|B) = 0$$

بما أنّ A و B حدثين متنافيين فإنّ: $A \cap B = \phi$ وبالتالي يكون $P(A \cap B) = 0$ ومنه نجد أنّ:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0$$

د) – إذا كان A و B حدثان مستقلان وغير متنافيين فإنّ:

$$P(A|B)=P(A) \quad , \quad P(B|A)=P(B) \qquad \textbf{(2.15)}$$

الاثبات:

ليكن A و B حدثين مستقلين وغير متنافيين عندئذٍ من تعريف الاحتمال الشرطي في العلاقة (2.10) نجد أنّ:

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

وبالاستفادة من العلاقة (2.2) نجد أنّ:

$$P(A|B)=\frac{P(A)P(B)}{P(B)}$$

وبما أنّ $P(B)\neq 0$ لأن الحدث B قد وقع فعلاً فإنّ:

$$P(A|B)=P(A)$$

بنفس الطريقة نثبت أنّ: $P(B|A)=P(B)$

ممّا سبق يمكننا صياغة التعريف الرياضي التالي للحوادث المستقلة:

2-6-2- تعريف الحوادث المستقلة رياضياً:

نقول عن الحدثين غير المتنافيين A, B أنّهما حدثان مستقلان إذا وفقط إذا كان:

$$P(A|B)=P(A) \quad , \quad P(B|A)=P(B)$$

أو كان:

$$P(A\cap B)=P(A).P(B) \qquad \textbf{(2.16)}$$

كما وجدنا سابقاً في الخاصية (2.2)

وبشكل عام نقول عن الأحداث غير المتنافية $A_1,A_2,\ldots,A_n$ أنّها أحداث مستقلة إذا اوفقط إذا كان احتمال تقاطعها يساوي جداء احتمالاتها أي أنّ:

$$P(A_1\cap A_2\cap\ldots\cap A_n)=P(A_1).P(A_2)\ldots P(A_n) \qquad \textbf{(2.17)}$$

وكانت مستقلة من أجل أي مجموعة جزئية من الصف السابق.

مثال (16): صندوق فيه 6 كرات، 2 بيضاء و 4 زرقاء. سحبنا عشوائياً كرتين. ما احتمال أن تكون الكرتان المسحوبتان زرقاوتين في الحالتين التاليتين:

1- إذا تمت عملية السحب على التتالي (دون إعادة)
2- إذا تمت عملية السحب بحيث تعاد الكرة المسحوبة للصندوق.

الحل: ليكن A الحدث الذي يقع إذا وفقط إذا كانت الكرة المسحوبة أولاً زرقاء. و B الحدث الذي يقع إذا وفقط إذا كانت الكرة المسحوبة ثانياً زرقاء. عندئذٍ يكون الحدث المطلوب هو $(A \cap B)$ حيث نميز الحالتين التاليتين:

1- إذا تمت عملية السحب بدون إعادة فإنّ:

$$P(A \cap B) = P(A).P(B|A)$$

حيث نلاحظ أنّ:

$$P(A) = \frac{4}{6} = \frac{2}{3}$$

$$P(B|A) = \frac{3}{5}$$

ومنه:

$$P(A \cap B) = \frac{2}{3}.\frac{3}{5} = \frac{2}{5} = 0.4$$

2- أمّا إذا تمت عملية السحب بحيث تُعاد الكرة المسحوبة للصندوق فإنّ الحدثين $A,\ B$ في هذه الحالة يكونا مستقلين ولهما نفس فرص الوقوع حيث نجد أنّ:

$$P(A) = \frac{4}{6} = \frac{2}{3}$$

$$P(B|A) = P(B) = \frac{4}{6} = \frac{2}{3}$$

وبالتالي فإنّ:

$$P(A \cap B) = P(A).P(B) = \frac{2}{3}.\frac{2}{3} = \frac{4}{9} \cong 0.444$$

مثال (17): صندوق فيه 25 مصباحاً كهربائياً 5 منها تالف. سحبنا عشوائياً ثلاثة مصابيح. ما احتمال أن تكون المصابيح المسحوبة جميعها تالفة في الحالتين التاليتين:

1- إذّا تمت عملية السحب على التتالي (دون إعادة)
2- إذّا تمت عملية السحب بحيث يُعاد المصباح المسحوب للصندوق.

الحل: ليكن A الحدث الذي يقع إذّا كان المصباح المسحوب أولاً تالف، و B الحدث الذي يقع إذا كان المصباح المسحوب ثانياً تالف و C الحدث الذي يقع إذّا كان المصباح المسحوب ثالثاً تالف. عندئذٍ يكون الحدث المطلوب هو $A \cap B \cap C$ حيث نميّز الحالتين التاليتين:

1- إذا تمت عملية السحب بدون إعادة فإنّ:

$$P(A\cap B\cap C)=P(A)P(B|A)P(C|(A\cap B))$$

حيث نلاحظ أنّ:

$$P(A)=\frac{5}{25}=\frac{1}{5}$$
$$P(B|A)=\frac{4}{24}=\frac{1}{6}$$
$$P(C|(A\cap B))=\frac{3}{23}$$

ومنه:

$$P(A\cap B\cap C)=(\frac{1}{5})(\frac{1}{6})(\frac{3}{23})=\frac{1}{230}\cong 0.004$$

2- إذا تمت عملية السحب بحيث يُعاد المصباح المسحوب للصندوق فإنّ الأحداث A, B, C في هذه الحالة تكون مستقلة ومتساوية فرص الوقوع حيث يكون:

$$P(A)=\frac{5}{25}=\frac{1}{5} \quad , \quad P(B)=\frac{5}{25}=\frac{1}{5} \quad , \quad P(C)=\frac{5}{25}=\frac{1}{5}$$

وبالتالي فإنّ:
$$P(A\cap B\cap C)=P(A)P(B)P(C)=\frac{1}{5}.\frac{1}{5}.\frac{1}{5}=\frac{1}{125}=0.008$$

مثال (18): لدى شخص جهاز كهربائي حسّاس يحتاج إلى منظم كهربائي. فإذا علمت أنّ احتمال شراء الشخص للمنظم هذا العام هو 0.70، واحتمال أن يتعطل الجهاز إذا وصل معه منظم هو 0.20، وأنّ احتمال تعطل الجهاز في حال عدم وصله بمنظم هو 0.85. ما احتمال أن يتعطل الجهاز هذا العام.

الحل: ليكن A الحدث الذي يقع إذا وفقط إذا تعطل الجهاز هذا العام و B الحدث الذي يقع إذا وفقط إذا اشترى الشخص منظم هذا العام أيضاً. عندئذٍ يكون الحدث المطلوب حساب احتماله هو تعطل الجهاز هذا العام سواءً اشترى الشخص منظماً أم لم يشتر منظماً أي الحدث $A\cap(B\cup B')$ حيث نجد أنّ:

$$P[A\cap(B\cup B')]=P[(A\cap B)\cup(A\cap B')]=P(A\cap B)+P(A\cap B')$$

وذلك بملاحظة أنّ الحدثين $A\cap B$, $A\cap B'$ متنافيان. ومنه نجد أنّ:

$$P[A\cap(B\cup B')]=P(B)P(A|B)+P(B')P(A|B')$$

$$= (0.70)(0.20) + (0.30)(0.85) = 0.395$$

مثال (19): آلة للنسيج تنتج قماشاً يحتوي على نسبة معينة من القطن. وجد المهندس المختص أنّ الآلة بحاجة لصيانة وإلاّ ستنتج قماشاً يحتوي على نسبة من القطن أكبر من النسبة المطلوبة. فإذا كان احتمال موافقة الإدارة على صيانة الآلة هذا العام هو 0.60 واحتمال أن تعطي الآلة قماشاً بنفس النسبة المطلوبة من القطن بدون صيانتها هو 0.40، أمّا إذا تم صيانة الآلة فاحتمال انتاجها لقماش يحتوي على النسبة المطلوبة من القطن هو 0.95. ما احتمال انتاج الآلة للقماش ضمن النسبة المطلوبة من القطن هذا العام؟

الحل: ليكن A الحدث الذي يقع إذا وفقط إذا أنتجت الآلة قماشاً يحتوي على النسبة المطلوبة من القطن هذا العام، و B الحدث الذي يقع إذا وفقط إذا وافقت الإدارة على صيانة الآلة هذا العام أيضاً. عندئذٍ يكون الحدث المطلوب حساب احتماله هو: $A \cap (B \cup B')$
حيث نجد أنّ:

$$P[A \cap (B \cup B')] = P[(A \cap B) \cup (A \cap B')] = P(A \cap B) + P(A \cap B')$$

وذلك بملاحظة أن الحدثين: $A \cap B$ و $A \cap B'$ متنافيان. ومنه نجد أنّ:

$$P[A \cap (B \cup B')] = P(B)P(A|B) + P(B')P(A|B')$$
$$= (0.60)(0.95) + (0.40)(0.40) = 0.73$$

2-7 – قانون بايز (Bays Rule):

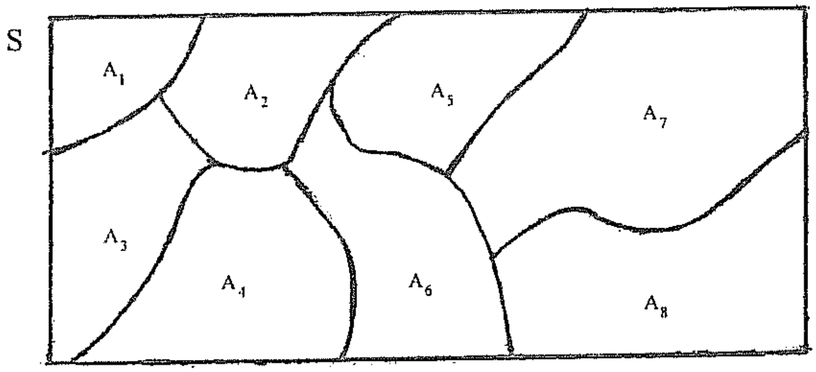
لنوضح أولاً مفهوم التجزئة لفضاء التجربة S.

تعريف (6): إذا كانت $(A_1, A_2, \ldots, A_n, \ldots)$ متتالية منتهية أو غير منتهية من الأحداث في الفضاء الاحتمالي (S, ξ, P) فإننا نقول عن هذه المتتالية أنّها تشكل تجزئة للحدث الشامل (أو الأكيد) S، إذا كانت تغطي S من جهة، ومتنافية مثنى مثنى من جهة أخرى. أي تكون المتتالية $(A_1, A_2, \ldots, A_n, \ldots)$ تجزئة لـ S إذا تحقق الشرطان التاليان:

أ) $S = \bigcup_{i \geq 1} A_i = A_1 \cup A_2 \ldots \cup A_n \cup \ldots$

ب) $i \neq j \Rightarrow A_i \cap A_j = \phi$

إنّ الشكل (2-7) يمثل تجزئة للحدث الشامل S إلى ثمانية أحداث.



الشكل (2-7)

2-7-1 قانون بايز:

إذا كانت الأحداث ... , $A_1, A_2, \ldots, A_n$ تجزئة للحدث الشامل S، وكان B حـــدثاً مرتبطـــاً بالتجربة ذاتها فإنّ:

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_{j\geq 1}P(A_j)P(B|A_j)} \qquad \textbf{(2.18)}$$

وفي حال كون التجزئة للحدث الشامل S منتهية أي إذا كانت $A_1, A_2, \ldots, A_n$ تجزئة لـــ S وكان B حدثاً مرتبطاً بالتجربة ذاتها عندئذٍ يصبح قانون بايز بالشكل:

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n}P(A_j)P(B|A_j)} \qquad \textbf{(2.19)}$$

وذلك لأنه إذا كانت $A_1, A_2, \ldots, A_n$ تجزئة لـــ S و B حدثاً ما مرتبط بنفس التجربة فإنـــه يمكننا كتابة ما يلي:

$$P(B\cap A_i)=P(A_i\cap B)$$

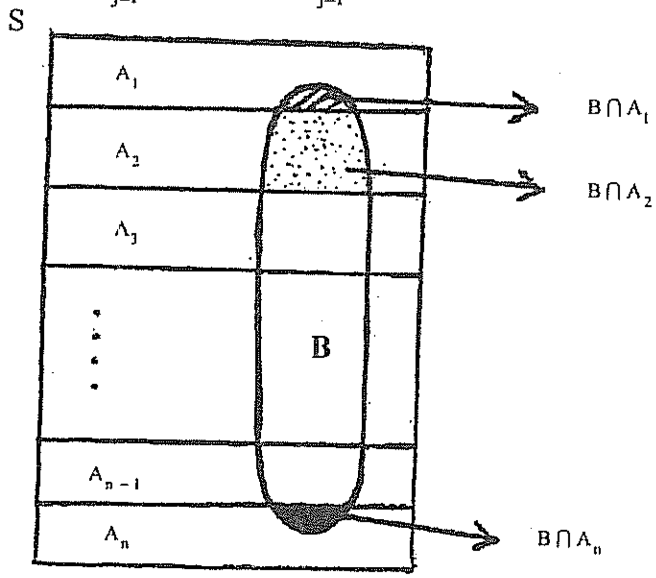
وذلك بملاحظة الشكل (2-8) وبالتالي فإنّ:

$$P(B)P(A_i|B)=P(A_i)P(B|A_i)$$

ومنه:

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{P(B)}=\frac{P(A_i)P(B|A_i)}{P(\bigcup_{j=1}^{n}(B\cap A_j))}$$

$$=\frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n}P(B\cap A_j)}=\frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n}P(A_j)P(B|A_j)}$$



الشكل (2-8)

مثال (20): لدينا ثلاثة صناديق يحوي كل منها كرات بيضاء وسوداء وحمراء. ولنفترض أنّ توزيع الكرات في هذه الصناديق ممثل بالجدول التالي:

اللون / الصندوق	أسود B	أبيض W	أحمر R
1	2	3	5
2	5	10	5
3	6	2	1

أخترنا اتفاقاً، صندوقاً من هذه الصناديق (دون أن نراه) ثم سحبنا منه كرة.

أ- ما هو احتمال أن تكون هذه الكرة سوداء.

ب- إذا علمت أن الكرة المسحوبة سوداء فما هو احتمال أن يكون السحب قد تم من الصندوق الأول؟

الحل: لنرمز بـ A_i للحدث "جرى السحب من الصندوق رقم i"، $i=1,2,3$ وليكن B الحدث "الكرة المسحوبة سوداء". إنّ من الواضح أن الأحداث A_i تشكّل تجزئة للحدث الشامل (أو الأكيد).

أ- إنّ المطلوب حساب $P(B)$ ولكن:

$$P(B)=P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+P(A_3)P(B|A_3)$$

$$P(B)=\frac{1}{3}.\frac{2}{10}+\frac{1}{3}.\frac{5}{20}+\frac{1}{3}.\frac{6}{9}\cong 0.372$$

ب- إنّ المطلوب حساب $P(A_1|B)$ حيث نجد أنّ:

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{P(B)}=\frac{P(A_1)P(B|A_1)}{\sum_{j=1}^{3}P(A_j)P(B|A_j)}$$

$$=\frac{(\frac{1}{3})(\frac{2}{10})}{0.372}\cong 0.179$$

مثال (21): ثلاث آلات M_1, M_2, M_3 تنتج 40% ، 35% ، 25% من انتاج مصنع وذلك على الترتيب. وجد أن 2٪ و 4٪ و 5٪ من إنتاج هذه الآلات رديء وذلك على الترتيب أيضاً. اخترنا سلعة من إنتاج هذا المصنع فكانت رديئة، فما هو احتمال كون هذه السلعة من انتاج الآلة M_1؟

الحل: لنرمز بـ A_i للحدث الذي يقع إذا كانت السلعة من إنتاج الآلة M_i، $i=1,2,3$ وبالرمز B للحدث "السلعة المسحوبة رديئة". من الواضح أن الأحداث A_i تشكل تجزئة للحدث الشامل وأنّ المطلوب حسابه هو: $P(A_1|B)$. حيث لدينا:

$$P(A_1)=0.25\ ,\ P(A_2)=0.35\ ,\ P(A_3)=0.40$$

$$P(B|A_1)=0.05\ ,\ P(B|A_2)=0.04\ ,\ P(B|A_3)=0.02$$

$$P(A_1|B) \frac{P(A_1)P(B|A_1)}{\sum_{j=1}^{3} P(A_j)P(B|A_j)} = \frac{(0.25)(0.05)}{(0.25)(0.05)+(0.35)(0.04)+(0.40)(0.02)}$$

$$= \frac{0.0125}{0.0345} = \frac{125}{345} \cong 0.362$$

2-8 طرائق العدّ:

وجدنا أنه لحل مسألة إحصائية معينة أننا بحاجة إلى معرفة فضاء التجربة حيث يكون معرفة عدد عناصر هذا الفضاء شاقاً في بعض الأحيان. لذلك سنستعرض بعض المبادئ من طرائق العد التي تساعدنا على معرفة عدد عناصر فضاء التجربة وبالتالي في حساب احتمال حدث ما.

2-8-1 المبدأ الأساسي للعد: Fundamental principle of counting

إذا أمكن إجراء عملية ما بـ n_1 طريقة مختلفة، وإذا تلت هذه العملية عملية أخرى يمكن إجراؤها بـ n_2 طريقة مختلفة، وإذا تلت العملية الثانية عملية ثالثة يمكن إجراؤها بـ n_3 طريقة مختلفة، وهكذا... وإذا تلت العملية (k-1) عملية أخرى يمكن إجراؤها بـ n_k طريقة مختلفة، فإنّ عدد الطرق التي يمكن إجراء هذه العمليات بها على الترتيب هو حاصل الضرب:

$$n_1.n_2....n_k \qquad \textbf{(2.20)}$$

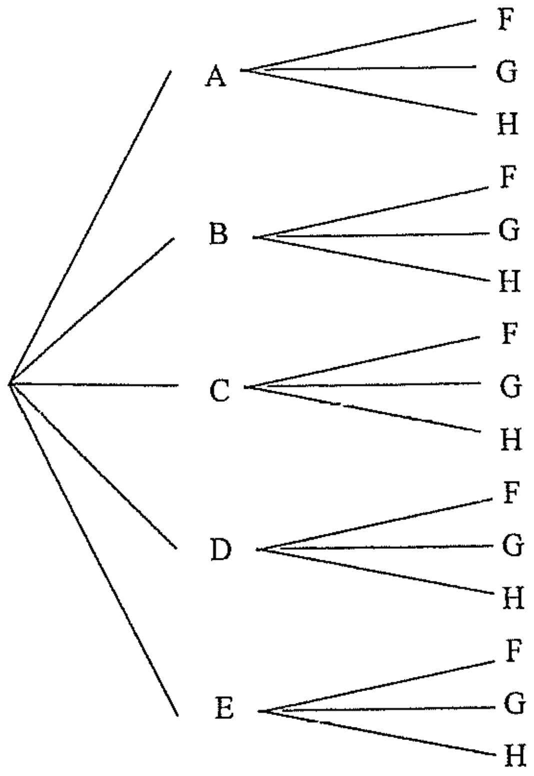
مثال (22): إذا كانت $A_1, A_2, ..., A_k$ مجموعات مختلفة عدد عناصرها على الترتيب: $n_1.n_2....n_k$. بكم طريقة يمكن اختيار عنصر واحد من كل مجموعة.

الحل: إنّ عدد الطرق الممكنة لاختيار عنصر واحد من كل مجموعة هو:

$$n_1.n_2.n_3....n_k$$

مثال (23): أعلنت شركة للصناعات النسيجية عن وظيفتين شاغرتين، فتقدم 3 رجال و 5 سيدات لملء هاتين الوظيفتين. بكم طريقة يمكن تعيين رجل وسيدة في هاتين الوظيفتين.

الحل: إذا رمزنا للسيدات المتقدمات للوظيفة بالرموز A, B, C, D, E وللرجال المتقدمين لهذه الوظيفة بالرموز F, G, H فإنه يمكننا تمثيل طرق الاختيار بالشجرة التالية:



بالاستفادة من المبدأ الأساسي للعد نجد أن هناك:

$$5 \times 3 = 15$$

طريقة مختلفة لاختيار رجل وسيدة من عدد المرشحين الكلي.

مثال (24): أعلنت مؤسسة الكهرباء في إحدى المدن عن وجود وظيفتين شاغرتين لرئيس قسم الصيانة ومعاونه، فتقدم للوظيفة الأولى 6 مهندسين اختصاص طاقة كهربائية. وتقدم للوظيفة الثانية 4 من خريجي المعهد الهندسي للكهرباء. كم عدد الطرق الممكنة لملء هاتين الوظيفتين الشاغرتين؟

الحل: نلاحظ أنه يوجد $n_1=6$ طرق لشغل وظيفة رئيس قسم الصيانة، ومع كل طريقة من هذه الطرائق يوجد $n_2=4$ طرق لشغل وظيفة معاون رئيس قسم الصيانة. وبالتالي فإنه يوجد $n_1.n_2 = 6\times 4 = 24$ طريقة ممكنة لملء هاتين الوظيفتين الشاغرتين.

2-8-2 التباديل Permutations:

في الكثير من الحالات والتجارب الاحصائية نحاول حساب أو إيجاد عدد التراتيب أو التباديل الممكن تكوينها من عدد من العناصر المختلفة وغير المتشابهة مثل ترتيب الألوان أو ترتيب الشخصيات ...إلخ. فمثلاً إذا كان لدينا ثلاثة عناصر مختلفة مثل A, B, C فإن عدد التراتيب أو التباديل الممكنة لهذه العناصر هو ستة. حيث تكون الحالات الممكنة لترتيب هذه العناصر هي:

ABC, ACB, BAC, BCA, CAB,CBA

وبصورة عامة نجد أنّ تعداد التراتيب أو التباديل الممكن إجراؤها على n من العناصر المختلفة هو :

$$P(n,n) = n! = n(n-1)(n-2)...3.2.1 \qquad (2.21)$$

حيث يدعى $P(n,n)$ عدد تباديل n عنصر مختلف.

مثال (25): بكم طريقة يمكن ترتيب 4 حواسيب جديدة مختلفة المنشأ في أحد مخابر الحواسيب بكلية الهندسة؟

الحل: يمكن ترتيب الحاسب الأول بأربع طرق والحاسب الثاني بثلاث طرق والحاسب الثالث بطريقتين وأخيراً الحاسب الرابع بطريقة واحدة. وبالتالي فإنّ عدد الطرق يكون:

$$P(4,4) = 4! = 4.3.2.1 = 24$$

2-8-3 التراتيب (Arrangements):

إذا كان المطلوب هو حساب عدد التراتيب لمجموعة k من العناصر والمأخوذة من n من العناصر المختلفة. وبعبارة أخرى ترتيب n من العناصر المختلفة على k من المراحل أو k من المواقع مثل ترتيب n من الأشخاص على k من الكراسي. وبشكل عام لو رمزنا بـ $P(n,k)$ لعدد تراتيب k عنصر مأخوذة معاً من n عنصر فإنّ:

$$P(n,k) = n(n-1)(n-2)...(n-k+1)$$

$$P(n,k)=\frac{n!}{(n-k)!} \qquad (2.22)$$

مثال (26): أعلنت شركة عن وظيفتين شاغرتين مختلفتين، فتقدم أربعة أشخاص. بكم طريقة يمكن اختيار اثنين منهم؟

الحل: إذا رمزنا للأشخاص الأربعة بالرموز A, B, C, D فإنّ الحالات الممكنة لترتيب شخصين منهم هي:

AB, AC, AD
BC, BD, BA
CD, CA, CB
DA, DB, DC

وبالتالي فإنّ عدد الطرق يكون:

$$P(4,2)=\frac{4!}{(4-2)!}=\frac{4!}{2!}=12$$

2-8-4 التوافيق (Combinations):

إذا كان المطلوب هو إيجاد عدد الطرق الممكنة لاختيار k عنصر من n من العناصر. أيّ أننا نريد في كل مرة اختيار k عنصر بغض النظر عن الترتيب بين هذه العناصر. وهذا يعني أننا نريد تجميعاً أو توفيقاً للعناصر فقط بدون اعتبار التبادل بين مواقع العناصر من جملة هذه التجمعات.

وبشكل عام لو رمزنا بـ $C(n,k)$ أو C_n^k أو $\binom{n}{k}$ لعدد توافيق k عنصر ماخوذة معاً من n عنصر فإنّ:

$$C_n^k=\frac{n!}{k!(n-k)!} \qquad (2.23)$$

ملاحظة ②: من التعريف السابق وبملاحظة العلاقة (2.22) نستنتج أنّ:

$$C_n^k=\frac{P(n,k)}{k!}$$

أي أنّ عدد توافيق k عنصر مأخوذة معاً من n عنصر يكون أصغر من عدد التراتيب للعناصر k.

نتيجة (3): من العلاقة (2.23) نجد أنّ: $C_n^n = 1$ ، $C_n^1 = n$ ، $C_n^0 = 1$ حيث يمكن إثبات ذلك بسهولة.

مثال (27): لتكن لدينا العناصر a, b, c, d والمطلوب:

أ- كم ثنائية مرتبة يمكن تشكيلها من هذه العناصر، مع عدم تكرار العنصر الواحد في الثنائية الواحدة؟

ب- بكم طريقة يمكن تشكيل مجموعات ذات عنصرين فقط من هذه العناصر؟

الحل:

أ- إنّ الحالات الممكنة للثنائيات المرتبة المطلوبة هي:

(a,b), (a,c), (a,d), (b,a), (b,c), (b,d), (c,a), (c,b), (c,d), (d,a), (d,b), (d,c)

حيث نلاحظ أنّ عدد هذه الثنائيات هو 12 وهو عبارة عن ترتيب عنصرين مأخوذين معاً من أربع عناصر أيّ أنّ:

$$P(4,2) = \frac{4!}{(4-2)!} = 12$$

ب- إنّ المجموعات الممكنة هي:

$$\{a,b\}, \{a,c\}, \{a,d\}, \{b,c\}, \{b,d\}, \{c,d\}$$

وعددها يساوي 6، وهو عبارة عن توافيق عنصرين مأخوذين معاً من أربعة عناصر أي أنّ:

$$C_4^2 = \frac{4!}{2!(4-2)!} = \frac{12}{2} = 6$$

مثال (28): ما عدد اللجان ذات الأعضاء الثلاثة التي يمكن تكوينها من خمسة أشخاص.

الحل: عند تكوين لجنة من ثلاثة أعضاء لا يهمنا الترتيب للأشخاص وبالتالي فإنّ عدد اللجان يكون:

$$C_5^3 = \frac{5!}{3!\ 2!} = 10$$

مثال (29): ما هو عدد الخلطات رباعية الألوان التي يمكن تكوينها من سبعة ألوان مختلفة؟

الحل: عند تكوين هذه الخلطات لا يراعى ترتيب الألوان في الخلطة الواحدة، لذا فـإنّ عـدد الخلطات الممكن هو:

$$C_7^4 = \frac{7!}{4!\ 3!} = 35$$

2-9 تمارين عامة:

1- ألقينا قطعة نقود، فإذا كان الناتج «صورة» أعدنا إلقاءها ثانية، أمّا إذا كان «كتابة» فإننا نلقي حجر النرد. أكتب فضاء التجربة.

2- أربعة طلاب انتخبوا عشوائياً من صف ما، فإذا رمزنا بـ M للطالب و بـ F للطالبة، أكتب فضاء التجربة S_1. ثم أكتب فضاء تجربة آخر S_2 عناصره تمثل عدد الإناث المنتخبة.

3- إذا كانت $S=\{1,2,3\}$ أثبت أنّ الصف $\xi=\{\phi,\{3\},\{1,2\},S\}$ هو جبر وجبر تام على S.

4- إذا كانت $S=\{a,b,c,d\}$ بين أنّ الصف $\zeta=\{\{b\},S,\{c,d\}\}$ ليس جبراً على S.

5- إذا كان الحدثان A,B معرفان على فضاء تجربة إحصائية عشوائية. فعبّر عن الرموز التالية بعبارات واضحة.

$$A',B',\ A\cup B,\ A\cap B,\ A'\cap B',\ A'\cup B',\ (A\cap B)',\ (A\cup B)'$$

6- إذا كانت الحوادث A,B,C معرّفة على فضاء تجربة عشوائية. فعبّر عن الحوادث التالية رياضياً:

أ- وقوع الحدث A مع عدم وقوع كلاً من الحدثين C , B.

ب- وقوع الحدث A مع عدم وقوع أحد الحدثين C , B.

جـ - وقوع الحوادث الثلاثة معاً.

د - عدم وقوع أي من الحوادث الثلاثة.

هـ - وقوع أحد الحوادث الثلاثة على الأقل.

7- استخدم مخططات فن لإثبات العلاقات التالية بين الأحداث المذكورة:

أ- $A = A\cup (A\cap B)$

ب- $A = (A\cap B)\cup(A\cap B')$

جـ - $(A\cap B)' = A'\cup B'$

د - $(A\cup B)' = A'\cap B'$

هـ - $(A \cup B) = (A \cap B) \cup (A \cap B') \cup (A' \cap B)$

و - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

8- ألقي حجري نرد، الأول أخضر والثاني أبيض، وسجلنا النتائج والمطلوب:

أ - أكتب فضاء التجربة.

ب- أكتب عناصر الحدث A الذي يقع إذا كان المجموع للرقمين الظاهرين على الحجرين أقل من 6 ثم أحسب احتمال A.

جـ - أحسب احتمال الحدث B الذي يقع إذا ظهر الرقم 2 على الحجر الأخضر.

د - أحسب احتمال الحدث C الذي يقع إذا ظهر الرقم 4 على أحد حجري النرد.

هـ - أحسب احتمال الحدث $B \cap C$

و - أحسب احتمال الحدث $A \cup B \cup C$

9- ألقي حجري نرد مختلفين والمطلوب:

1- أكتب فضاء التجربة.

2- إذا كان A الحدث الذي يقع إذا كان المجموع للرقمين الظاهرين هو 7 و B الحدث الذي يقع إذا كان المجموع لهذين الرقمين هو 9 فاحسب كلاً من الاحتمالات التالية:

$P(A), P(B), P(A \cap B),\ P(A)', P(B'), P(A \cup B), P(A \cup B'), P(A' \cup B), P(A' \cap B')$

10- تقدم رجلان وامرأتان بطلباتهم لملء وظيفتين مختلفتين شاغرتين في إحدى المؤسسات. اختار المدير عشوائياً شخصاً للوظيفة الأولى وآخر للوظيفة الثانية. والمطلوب:

أ- أكتب فضاء التجربة S.

ب- أحسب احتمال الحدث A الذي يقع إذا كان رجلاً قد شغل الوظيفة الأولى.

جـ - أحسب احتمال الحدث B الذي يقع إذا كانت وظيفة واحدة فقط قد شغلتها إمرأة.

د - أحسب احتمال الحدث C الذي يقع إذا كانت الوظيفتان قد شغلتا برجلين.

11- سحبت عشوائياً كرة من كيس يحتوي 3 كرات حمراء، 4 زرقاء، و 5 صفراء.

أ - ما احتمال أن لا تكون الكرة المسحوبة صفراء اللون.

ب- ما احتمال كون الكرة المسحوبة صفراء أو زرقاء.

12- كيس يحوي 4 كرات بيضاء و 3 سوداء، وكيس آخر يحوي 3 بيضاء و 5 سوداء. أختيرت كرة عشوائياً من الكيس الأول ووضعت في الكيس الثاني. سحبت بعد ذلك كرة من الكيس الثاني. ما احتمال أن تكون الكرة المسحوبة من الكيس الثاني سوداء؟

13- هل يمكن لحدثين A و B من فضاء احتمالي أن يحققا ما يلي:

$$P(A)=\frac{1}{8} \text{ ، } P(B)=\frac{4}{5} \text{ ، } P(A\cup B)=\frac{1}{2}$$ ولماذا؟

14- صندوق يحتوي على 5 كرات حمراء و 4 كرات بيضاء و 3 كرات سوداء سحبنا كرة من الصندوق بشكل عشوائي. أحسب الاحتمال في كل من الحالات التالية:

أ - إذا كانت الكرة المسحوبة حمراء.

ب- إذا كانت الكرة المسحوبة سوداء.

جـ - إذا كانت الكرة المسحوبة ليست حمراء.

د - إذا كانت الكرة المسحوبة ليست حمراء أو بيضاء.

هـ - إذا كانت الكرة المسحوبة بيضاء.

15- مائة بطاقة يانصيب مرقمة من 1 إلى 100، سحبت بطاقة عشوائياً. ما هو احتمال أن يكون رقم البطاقة المسحوبة:

أ - من مضاعفات الخمسة.

ب- من مضاعفات العدد 8.

جـ من مضاعفات العدد 5 أو 8.

د - لا يقبل القسمة على العدد 5.

16- إذا كان A و B حدثين مستقلين فأثبت أنّ الحدثين A' و B' مستقلان أيضاً.

17- إذا كان A و B حدثين معرّفين في فضاء تجربة ما. وكان:

$$P(A\cup B)=\frac{1}{2} \text{ ، } P(B)=\frac{1}{3} \text{ ، } P(A)=t$$

فأحسب قيمة t في الحالتين التاليتين:

أ – إذا كان الحدثان A , B متنافيين.

ب – إذا كان الحدثان A , B مستقلين.

18– صنعت قطعة نقود بحيث أن احتمال ظهور الصورة هو $P(H)=\frac{1}{3}$ واحتمال ظهور الكتابة $P(T)=\frac{2}{3}$ فإذا ألقيت هذه القطعة مرة واحدة وظهرت الصورة فإننا نختار عدداً عشوائياً من 1 إلى 11، أمّا إذا ظهرت الكتابة فإننا نختار عدداً عشوائياً من 1 إلى 7. فما هو احتمال أن يكون العدد المختار فردياً؟

19– تنوي أسرة قضاء إجازة نهاية أسبوع في إحدى الأماكن السياحة A ، B ، أو C فإذا كان احتمال سقوط المطر في A هو 0.6 وفي B هو 0.7 وفي C هو 0.5 وإذا اختارت الأسرة مكان الاجازة عشوائياً فأحسب:

أ – احتمال أن تكون الإجازة ممطرة.

ب – إذا كانت الإجازة ممطرة فما هو احتمال أن هذه الإجازة كانت في المكان B.

20– ثلاث قطع نقود الأولى C_1 متزنة والثانية C_2 والثالثة C_3 غير متزنين. فإذا كان احتمال ظهور الصورة على القطعة الثانية هو 0.4 واحتمال ظهورها على القطعة الثالثة هو 0.75:

أ – أوجد احتمال الحصول على صورة واحدة إذا علمت أن اختيار أي قطعة يتم بصورة عشوائية.

ب – إذا كانت نتيجة الرمية كتابة فما هو احتمال أن تكون القطعة المقذوفة متزنة.

21– سلمى طالبة في السنة الأخيرة، احتمال أن تتخرج هذا العام هو 0.80 واحتمال أن تتزوج في صيف هذا العام إذا تخرجت هو 0.90، أمّا إذا لم تتخرج فاحتمال زواجها في ذلك الوقت هو 0.30. ماهو احتمال زواج سلمى صيف هذا العام؟

22– أراد مجموعة من الطلاب القيام برحلة يوم الجمعة القادم. أذاعت دائرة الأرصاد الجوية أن احتمال سقوط المطر يوم الجمعة القادم هو 25%. فإذا علمت أنه عندما يكون الجو

ممطراً فإن الطلاب سيعتبرون الرحلة جيدة باحتمال 30% وإذا كان صحواً فإنهم سيعتبرون الرحلة جيدة باحتمال 90%. ما احتمال أن تكون الرحلة جيدة؟

23- قرع الجرس في مدرسة ابتدائية فهرع تلاميذ الصف الرابع الذي يشتمل على ثلاثين طالباً إلى الانتظام في صف أحادي فإذا علمت أن سعيداً وأسعد تلميذان في هذا الصف.

أ – ما هو احتمال أن يكونا متجاورين.

ب – إذا علمت أنهما متجاوران فما هو احتمال أن يكون أسعد في آخر الصف.

24- ينتج أحد المصانع 300 بدلة رجالية و 900 بدلة نسائية كما ينتج مصنع آخر 500 بدلة رجالية و 500 بدلة نسائية فإذا أختيرت بدلة واحدة فقط من كل مصنع لقياس جودة الإنتاج، فما هو احتمال أن تكون إحدى البدلتين رجالية على الأقل.

25- إذا كان معدل إنتاج ثلاثة مصانع مختلفة للبطاريات هي 500، 700، 900 بالترتيب وكان عدد البطاريات التالفة في هذا الانتاج هو 20 ، 30 ، 60 على الترتيب حيث يُباع هذا الإنتاج في محل لبيع المفرق فما هو احتمال أن تكون البطارية المشتراة تالفة. ثم ما هو احتمال أن تكون البطارية المشتراة التالفة هي من إنتاج المصنع الأول؟

26- أثبت أنّ:

أ – $C_n^p = C_n^{n-p}$

ب – $C_n^p = C_{n-1}^p + C_{n-1}^{p-1}$

جـ – $C_n^0 + C_n^1 + \dots + C_n^n = 2^n$

27- أرادت شركة هندسية تشكيل لجنة مؤلفة من 5 مهندسين أثنان منهم من الشباب وثلاث من الفتيات. حيث تم ترشيح 5 شباب و 7 فتيات لهذه اللجنة. بكم طريقة يمكن تشكيل اللجنة إذا كان:

أ – بإمكان أي شاب أو فتاة دخول اللجنة.

ب – شاب معيّن يجب أن يكون ضمن اللجنة.

جـ – فتاتان معينتان يجب أن لا تكونا ضمن اللجنة.

28- إذا كان احتمال أن يعيش هيثم 20 عاماً آخر هو 60٪ واحتمال أن تعيش زوجة هيثم 20 عاماً أيضاً هو 80٪. فما هو احتمال أن يظل الاثنان على قيد الحياة 20 عاماً؟

29- صندوق يحوي 8 كرات حمراء و 3 خضراء و 9 زرقاء. سحبت 3 كرات عشوائياً. أوجد الاحتمالات التالية:

أ – أن تكون الكرات الثلاث زرقاء.

ب – أثنتان زرقاء وواحدة خضراء.

جـ – على الأقل كرة خضراء.

د – كرة من كل لون.

هـ – الأولى زرقاء والثانية حمراء والثالثة خضراء.

الفصل الثالث

المتغيرات العشوائية

3-1 – مقدمة.

3-2 – المتغير العشوائي.

3-3 – التوزيع الاحتمالي.

3-4 – التوزيع المشترك لمتغيرين عشوائيين.

3-5 – القيم المميزة للمتغيرات العشوائية.

3-6 – العزوم والدوال المولدة لها.

3-7 – تمارين عامة.

المتغيرات العشوائية
(Random Variables)

3-1 مقدمة:

درسنا في الفصول السابقة بعض الطرائق المستخدمة لوصف مجموعات من البيانات الإحصائية حيث كانت بعض هذه المجموعات ممثلة لمجتمع إحصائي محدود وبعضها الآخر ممثلة لعينة مأخوذة من مجتمع إحصائي غير محدود.

ثم ناقشنا نظرية الاحتمال، حيث استخدمنا بديهيات ومبرهنات الاحتمال في الإجابة عن العديد من الأسئلة المطروحة ذات الطبيعة المعقدة.

سنرى في فقرات قادمة أهمية نظرية الاحتمالات بالنسبة إلى تحليل البيانات، وبالتالي أهمية هذه النظرية في الحصول على استقراءات حول المجتمع المدروس معتمدين في ذلك على العملية المأخوذة من هذا المجتمع. وعند القيام بهذا العمل يجب علينا الاهتمام بموضوع الربط بين نظرية الاحتمال والإحصاء. وبخاصة سوف نوضح فكرة المتغير العشوائي ونبين أنّ معظم الدراسات الإحصائية هي في الحقيقة دراسات لخواص المتغيرات العشوائية وهكذا فإنّ استيعاب هذه الخواص ضروري لفهم الطرائق الإحصائية المذكورة في هذا الكتاب.

إذا كان لدينا تجربة عشوائية ما فإنّ هذه التجربة تقودنا إلى مجموعة من النتائج أو البيانات. إنّ هذه النتائج قد تكون كمية أو كيفية. فمثلاً عند رمي قطعتي نقود دفعة واحدة (أو رمي قطعة واحدة مرتين متتاليتين) فإنّ فضاء التجربة يكون: $S_1 = \{HH, HT, TH, TT\}$ وذلك إذا كان اهتمامنا هو النتيجة الظاهرة على القطعتين من حيث كونها صورة أو كتابة وفي حالة اهتمامنا بعدد الصور الظاهرة يصبح فضاء التجربة بالشكل: $S_2 = \{0,1,2\}$ حيث نلاحظ أنّ

البيانات الناتجة عن التجربة العشوائية في S_1 هي بيانات كيفية وفي S_2 هي بيانات كمية. ولكن نحتاج أحياناً لرد البيانات الكيفية إلى رقمية، حيث يُساعد ذلك الإحصائي على تسهيل الدراسة الإحصائية واستقراء النتائج. فيمكن أن نقرن كل عنصر من فضاء التجربة بعدد يعبر عن صفة معينة في العنصر. فمثلاً في مثالنا السابق يمكن أن نقرن كل عنصر من عناصر S_1 بعدد مرات ظهور الكتابة.

3-2 المتغير العشوائي (Random Variable):

تعريف (١):

المتغير العشوائي هو دالة من فضاء تجربة ما إلى مجموعة الأعداد الحقيقية R. حيث يقابل كل عنصر من فضاء التجربة عدداً حقيقياً هو قيمة المتغير العشوائي عن هذا العنصر.

يُرمز للمتغيرات العشوائية عادةً بأحرف لاتينية كبيرة مثل ...,X, Y, Z ويُرمز لقيم هذه المتغيرات العشوائية بأحرف لاتينية صغيرة مثل ...x, y, z .

فمثلاً في تجربة رمي قطعتي نقود دفعة واحدة، إذا كان X المتغير العشوائي الدال على عدد مرات ظهور الكتابة فإنّ X يعتبر دالة من فضاء التجربة S_1 (المبين في الفقرة السابقة) إلى مجموعة الأعداد الحقيقية R حيث يقرن المتغير X كل عنصر من S_1 بعدد مرات ظهور الكتابة في هذا العنصر. وبالتالي نجد أنّ:

$$X(HH) = 0 , X(HT) = X(TH) = 1 , X(TT) = 2$$

من التعريف السابق للمتغير العشوائي نجد أنّ الصورة العكسية لكل قيمة من قيم المتغير العشوائي X تمثل حدثاً. ففي مثال رمي قطعتي النقود المذكور أعلاه نجد أنّ:

$$X^{-1}(0) = \{HH\} , X^{-1}(1) = \{HT, TH\} , X^{-1}(2) = \{TT\}$$

وبالتالي لو أخذنا الحدث $E = \{HT, TH\}$ على سبيل المثال فإنّ:

$P(E) = \frac{2}{4} = \frac{1}{2}$ وكذلك نجد أنّ: $P(X = 1) = P(E) = \frac{1}{2}$ حيث نلاحظ هنا أن الحدثين المقابلين لقيمتين مختلفتين للمتغير العشوائي هما حدثان متنافيان.

مثال (1): عائلة عندها ثلاثة أولاد. أكتب فضاء التجربة S. ثم بفرض المتغير العشوائي X يدل على عدد الإناث، أذكر قيم هذا المتغير المقابلة لعناصر الفضاء S. ثم أحسب احتمال الحدث $X \le 2$.

الحل: إذا رمزنا بالرمز B للذكر والرمز G للأنثى فإنّ فضاء التجربة يكون:

$$S = \{BBB, BBG, BGB, BGG, GBB, GBG, GGB, GGG\}$$

ومنه فإنّ قيم المتغير العشوائي X تكون:

عناصر S	BBB	BBG	BGB	BGG	GBB	GBG	GGB	GGG
قيم X	0	1	1	2	1	2	2	3

وبالتالي فإنّ: $P(X \le 2) = \frac{7}{8} = 0.875$

كذلك يمكن حساب احتمال الحدث الأخير بطريقة أخرى، عن طريق الحدث المتمم، كما يلي:

$$P(X \le 2) = 1 - P(X = 3)$$
$$= 1 - \frac{1}{8} = 0.875$$

مثال (2): صندوق فيه سبعة مصابيح كهربائية ثلاثة منها تالف. سحبنا مصباحين عشوائياً، أذكر فضاء التجربة. ثم بفرض X المتغير العشوائي الدال على عدد المصابيح غير التالفة المسحوبة، أذكر القيم المقابلة. ثم أحسب احتمال أن نكون اخترنا مصباحاً واحداً على الأقل غير تالف.

الحل: إذا رمزنا بالرمز G للمصباح غير التالف والرمز B للمصباح التالف فإنّ فضاء التجربة يكون:

$$S = \{GG, GB, BG, BB\}$$

ومنه فإنّ قيم المتغير العشوائي X تكون:

عناصر S	GG	GB	BG	BB
قيم X	2	1	1	0

وبالتالي فإنّ احتمال أن نكون اخترنا مصباح واحد على الأقل غير تالف هو:

$$P(X \geq 1) = P(X = 1) + P(X = 2)$$

وبما أنّ:

$$P(X = 1) = \frac{C_4^1 C_3^1}{C_7^2} = \frac{(4)(3)}{21} = \frac{4}{7}$$

و:

$$P(X = 2) = \frac{C_4^2 C_3^0}{C_7^2} = \frac{(6)(1)}{21} = \frac{2}{7}$$

فإنّ:

$$P(X \geq 1) = \frac{4}{7} + \frac{2}{7} = \frac{6}{7} = 0.86$$

كذلك يمكن حساب الاحتمال الأخير بطريقة أخرى كما يلي:

$$P(X \geq 1) = 1 - P(X = 0)$$

$$= 1 - \frac{C_4^0 C_3^2}{C_7^2} = 1 - \frac{3}{21} = 0.86$$

3-3 التوزيع الاحتمالي (Probability Distribution):

يطلق عادةً على التوزيعات التي تَستَخدمُ الاحتمالات بدلاً من التكرارات أسم التوزيعات الاحتمالية. كما يطلق على هذه التوزيعات الاحتمالية أسم توزيعات احتمالية منقطعة أو منفصلة إذا كان فضاء التجربة العشوائية منقطعاً (أو منفصلاً) وبخلاف ذلك أي عندما يكون فضاء التجربة العشوائية متصلاً (أو مستمراً) فإنّ التوزيعات الاحتمالية تسمى بأسم التوزيعات الاحتمالية المستمرة. ويطلق على المتغير العشوائي المعرف على فضاء تجربة منفصل أسم المتغير العشوائي المنقطع (أو المنفصل). أمّا المتغير العشوائي المعرف على فضاء تجربة متصل فيطلق عليه أسم المتغير العشوائي المستمر(او المتصل). فمثلاً نلاحظ أنّ المتغيرين العشوائيين المعرّفين في المثالين (1 و 2) السابقين هو متغيران منقطعان لأنّ فضائي التجربة المعرّفين عليهما يمثلان فضاءين منفصلين.

3-3-1 التوزيع الاحتمالي المنقطع (أو المنفصل) (Discreate Probability Distribution) :

إنّ الدالة المعرفة على مجموعة قيم المتغير العشوائي المنقطع X، والتي تأخذ قيمها من مجموعة الأعداد الحقيقية R، والممثلة لاحتمال كل قيمة من قيم X تدعى بدالة الاحتمال المنقطع، أو دالة التوزيع الاحتمالي المنقطع ونرمز لها بالرمز f(x)

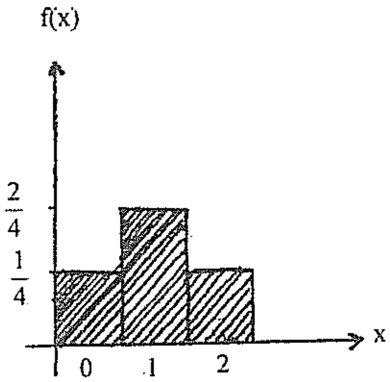
حيث $f(x) = P(X = x)$. كما يدعى الجدول الممثل لاحتمالات قيم المتغيــر العشــوائي X بجدول التوزيع الاحتمالي المنقطع.

مثال (3): في تجربة رمي قطعتي نقود متوازنتين دفعة واحدة، ليكن X المتغيــر العشــوائي الدال على عدد الصور الظاهرة. أوجد جدول التوزيع الاحتمالي لـ X.

الحل: إنّ فضاء التجربة في هذا المثال هو: $S = \{HH, HT, TH, TT)$
حيث يدل الرمز H على ظهور الصورة والرمز T على ظهور الكتابة. ومنه فــإنّ المتغيــر العشوائي X يأخذ القيم صفر أو 1 أو 2 وبالتالي نجد أنّ جدول التوزيــع الاحتمــالي لهــذا المتحول هو:

قيم X	0	1	2
$f(x) = P(X = x)$	$1/4$	$2/4$	$1/4$

يمكن تمثيل التوزيع الاحتمالي السابق بمدرج، كما فعلنا في حالــة مــدرج التكــرار النســبي للبيانات. حيث نمثل هذا التوزيع بمستطيلات متساوية العرض. عرض كــل منهــا الواحــد، ونجعل قيم المتغير العشوائي مراكز لقواعد هذه المستطيلات. أمّا أطوال هذه المستطيلات فكل منها يمثل الاحتمال لقيمة المتغير العشوائي الموافقة كما في الشكل (1-3) التالي:



الشكل (1-3)

إنّ مساحة كل مستطيل من المستطيلات السابقة تساوي احتمال مركز قاعدة هذا المستطيل.
كما أنّ مجموع مساحات المستطيلات كلها يساوي الواحد لأنّ:

$$\sum_{x=0}^{2} f(x)=\sum_{x=0}^{2} P(X=x)=\frac{1}{4}+\frac{2}{4}+\frac{1}{4}=\frac{4}{4}=1$$

مثال (4): صندوق فيه عشرة مصابيح كهربائية، اثنان منها تالف. اشترى شخص مصباحين أختيرا عشوائياً. بفرض المتغير العشوائي X يدل على عدد المصابيح التالفة التي أشتراها الشخص. أوجد جدول التوزيع الاحتمالي لـ X

الحل: إنّ فضاء التجربة في هذا المثال هو: $S = \{GG, GB, BG, BB)$
حيث يدل الرمز B على المصباح التالف والرمز G على المصباح غير التالف. ومنه فإنّ المتغير العشوائي X يأخذ القيم صفر أو 1 أو 2. والتي تحسب احتمالاتها كما يلي:

$$f(0) = P(X = 0) = \frac{C_8^2 C_2^0}{C_{10}^2} = \frac{28}{45}$$

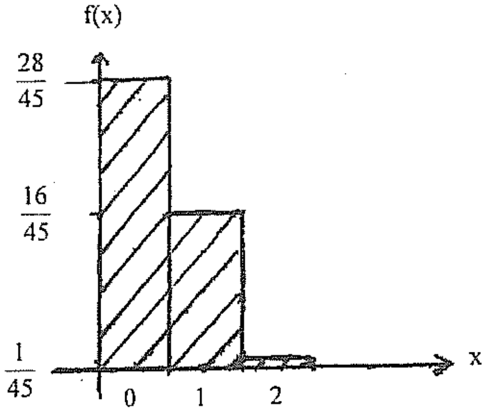
$$f(1) = P(X = 1) = \frac{C_8^1 C_2^1}{C_{10}^2} = \frac{16}{45}$$

$$f(2) = P(X = 2) = \frac{C_8^0 C_2^2}{C_{10}^2} = \frac{1}{45}$$

وبالتالي فإنّ جدول التوزيع الاحتمالي المنقطع للمتغير العشوائي X هو:

قيم X	0	1	2
$f(x) = P(X = x)$	$28/45$	$16/45$	$1/45$

يمكن تمثيل التوزيع الاحتمالي السابق بالمدرج المبين في الشكل (2-3) كما فعلنا في المثال (3):



الشكل (3-2)

حيث نلاحظ أنّ:

$$\sum_{x=0}^{2} f(x)=\sum_{x=0}^{2} P(X=x)=\frac{28}{45}+\frac{16}{45}+\frac{1}{45}=\frac{45}{45}=1$$

وبشكل عام نستنتج أنه إذا كان X متحولاً عشوائياً منقطعاً فإن:

$$\sum_{x} f(x)=1 \qquad (3.1)$$

حيث $0 \le f(x) \le 1$ أياً كان x من قيم المتغير العشوائي X.

إنّ استخدامنا لمفهوم المساحة في تمثيل الاحتمال ضروري لاعتبارات قادمة سنراها في حالة المتغير العشوائي المستمر.

3-3-2 التوزيع الاحتمالي المستمر (أو المتصل) (Continuous Probability Distribution)

وجدنا سابقاً أنّ بعض التجارب العشوائية يمكن أن تعطي نتائج ممكنة غير منتهية وغير قابلة للعد كتجربة تحديد المسافة التي يمكن أن تقطعها سيارة معينة بعشرين ليتراً من البنزين حيث نلاحظ هنا أنّ النتائج الممكنة لهذه التجربة تشكل مجموعة غير منتهية وغير قابلة للعد أي يوجد لهذه التجربة العشوائية فضاء تجربة متصل. وكما وجدنا سابقاً أيضاً فإنّ المتغير العشوائي المعرف على مثل هذه الفضاءات يكون متغيراً عشوائياً مستمراً. حيث تكون قيم هذا المتغير المقابلة لعناصر فضاء التجربة، عبارة عن مجموعة جزئية من مجموعة الأعداد الحقيقية. ونلاحظ في المتغير العشوائي المستمر أنّ احتمال قيمة معينة من قيمه يساوي

الصفر، وهذا ما لم نصادفه في المتغير العشوائي المنقطع حيث يُعزى ذلك بشكل بديهي إلى أنه إذا سئلنا مثلاً في تجربة قياس المسافة التي تقطعها سيارة معينة بعشرين ليتراً من البنزين ما احتمال أن تقطع السيارة مسافة قدرها 213.271891 كم بالضبط؟ واضح أنه من المستحيل أن تقطع السيارة هذه المسافة المحددة وبهذه الدقة بعشرين ليتراً من البنزين. وبالتالي فإنّ هذا الاحتمال يكون صفراً.

وجدنا أنه يمكن تمثيل التوزيع الاحتمالي المنقطع بجدول وبالتالي بمدرج أو مضلع. أمّا في حالة المتغير العشوائي المستمر فإنه ليس بالإمكان تمثيل التوزيع الاحتمالي له بجدول، وبالتالي لا يمكن تمثيله بيانياً بمدرج أو مضلع. بل نجد أنه عندما يكون المتغير العشوائي مستمراً فإنّ التوزيع الاحتمالي لهذا المتغير سيمثل بمنحنٍ مستمر، وهذا المنحني قد يكون بالإمكان التعبير عنه بمعادلة. إنّ هذه المعادلة تكون عادةً تابعة للقيم العددية التي يأخذها المتغير العشوائي المستمر، فإذا كان X متغيراً عشوائياً مستمراً وكانت x إحدى قيمه فإنّ معادلة المنحني الممثل لهذا المتغير تكون $y = f(x)$. إنّ الدالة $f(x)$ تُدعى عادةً بدالة الكثافة. حيث يجب أن تحقق هذه الدالة شروطاً معينة كي تمثل دالة كثافة لمتغير عشوائي ما.

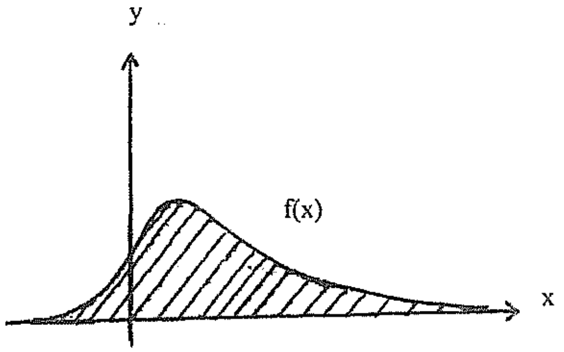
3-3-3 دالة الكثافة الاحتمالية (Probability density function):

ليكن X متغيراً عشوائياً مستمراً، عندئذٍ ندعو الدالة $y = f(x)$ المعرفة على مجموعة الأعداد الحقيقية أو على جزء منها، بدالة الكثافة الاحتمالية لهذا المتغير إذا كانت تحقق ما يلي:

(1) $f(x) \geq 0$ وذلك أياً كان x من قيم المتغير العشوائي X. أي أنّ الدالة $f(x)$ هي دالة غير سالبة.

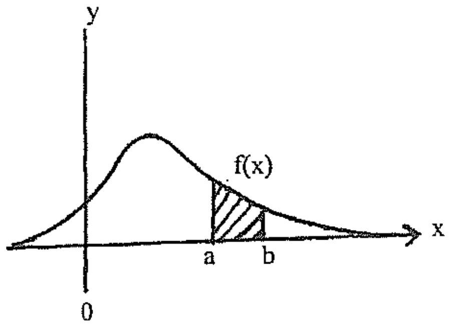
(2) $\int_{-\infty}^{\infty} f(x)dx=1$. أي أنّ المساحة المحددة بمنحني الدالة والمحور Ox تساوي الواحد.

ويمكن توضيح ذلك من خلال الشكل (3-3).



الشكل (3-3) مساحة المنطقة المظللة =1

من خلال التعريف السابق لدالة الكثافة الاحتمالية لمتغير عشوائي مستمر X نجد أنّ احتمال الحدث $a \le X \le b$ حيث a و b أي عددين حقيقيين من قيم المتغير X، سوف يعبر عنه بالمساحة المحدودة بمنحني دالة الكثافة والمحور ox والمستقيمين $x = a$, $x = b$ ، وذلك كما هو موضح بالشكل (3-4):



الشكل (3-4) مساحة المنطقة المظللة تمثل الاحتمال $P(a \le X \le b)$

كما نلاحظ أنّ: $P(a < X < b) = P(a \le X \le b)$

لأنّ

$$P(a \le X \le b) = P(a < X < b) + P(X = a) + P(X = b)$$

$$= P(a < X < b) + \int_a^a f(x)dx + \int_b^b f(x)dx$$

$$= P(a < X < b)$$

وذلك لأنّ: $\int_a^a f(x)dx = \int_b^b f(x)dx = 0$

ملاحظة ①: إذا كان X متغيراً عشوائياً منقطعاً فإنّ دالة الكثافة الاحتمالية لهذا المتغير هي نفسها دالة التوزيع الاحتمالي المنقطع $f(x) = P(X = x)$ التي عُرفت في الفقرة 1-3-3 حيث نلاحظ أنها تحقق الشرطين التاليين:

(1) $f(x) = P(X = x) \geq 0$ أياً كان x من قيم المتغير العشوائي X.

(2) $\sum_x f(x) = 1$.

ملاحظة ②: قد تكون دالة الكثافة الاحتمالية لمتغيرٍ عشوائي مستمر X معرفةً على مجالٍ ما [a,b] من مجموعة الأعداد الحقيقية R. في هذه الحالة يمكن اعتبار الدالة f(x) معرفة على كل R بحيث تأخذ القيمة صفر من أجل جميع قيم X التي لا تنتمي للمجال [a,b]. أي نعتبر دالة الكثافة بالشكل:

$$f(x) = \begin{cases} f(x) & , \quad x \in [a,b] \\ 0 & , \quad x \notin [a,b] \end{cases}$$

3-3-4 دالة التوزيع المجمعة (التراكمية):

تعود أهمية دالة التوزيع المجمعة إلى استخدامها في حساب الاحتمالات، حيث نلاحظ أنّ معظم الجداول الإحصائية المستخدمة بشكل واسعٍ لحساب الاحتمالات من أجل بعض المتغيرات العشوائية المستمرة الشهيرة، هي جداول لدوال التوزيع المجمعة لهذه المتغيرات.

فيما يلي سنعرّف دالة التوزيع المجمعة لمتغيرٍ عشوائي X سواءً كان منقطعاً أو مستمراً.

ليكن X متغيراً عشوائياً منقطعاً أو مستمراً، دالة كثافته الاحتمالية f(x)، عندئذ نُعرّف دالة التوزيع المجمعة للمتغير X والتي نرمز لها بالرمز F(x) بالعلاقة التالية:

$$F(x) = P(X \leq x) \qquad (3.2)$$

وبالتالي إذا كان X متغيراً عشوائياً منقطعاً فإنّ:

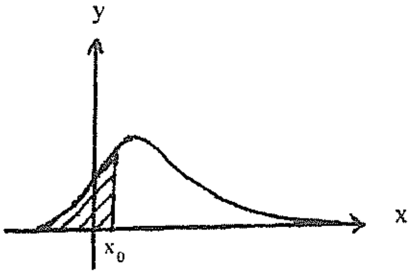
$$F(x_0) = P(X \leq x_0) = \sum_{x \leq x_0} f(x) = \sum_{x \leq x_0} P(X=x)$$

حيث x_0 قيمة ما للمتغير المنقطع X.

أمّا إذا كان X متغيراً عشوائياً مستمراً فإنّ:

$$F(x_0) = P(X \le x_0) = \int_{-\infty}^{x_0} f(x)dx$$

حيث x_0 قيمة ما للمتغير المستمر X. ويمكن توضيح ذلك بالشكل (3-5).



الشكل (3-5): $F(x_0) = P(X \le x_0)$

من العلاقة (3.2) نلاحظ أنّ دالة التوزيع المجمعة هي عبارة عن دالة في المتغير الحقيقي x وقيمها تتغير مع تغير x ، ففي الحالة التي يكون فيها المتغير العشوائي X مستمراً نلاحظ أنّ العلاقة الممثلة لدالة التوزيع المجمعة لهذا المتغير نحصل عليها من التكامل:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx \qquad \textbf{(3.3)}$$

كما يمكننا بالاستفادة من مفهوم دالة التوزيع المجمعة، حساب احتمال الحدث $P(x_1 < X \le x_2)$ بالشكل التالي:

$$P(x_1 < X \le x_2) = P(X \le x_2) - P(X \le x_1)$$

أو:

$$P(x_1 < X \le x_2) = F(x_2) - F(x_1) \qquad \textbf{(3.4)}$$

مع الملاحظة بأنّ الاحتمال السابق يجب أن يكون للحدث $x_1 < X \le x_2$ كما هو معطى تماماً من أجل المتغير العشوائي المنقطع. أمّا في حالة المتغير العشوائي المستمر فإنّ:

$$P(x_1 < X \le x_2) = P(x_1 \le X \le x_2) = P(x_1 \le X < x_2)$$

كما بينا سابقاً.

مثال (5): متغير عشوائي مستمر X دالة كثافته هي: $f(x) = \frac{2}{27}(1+x) \quad , \quad 2 \le x \le 5$

والمطلوب:

1- تأكد أنّ $f(x)$ هي دالة كثافة احتمالية.

2- أوجد دالة التوزيع المجمعة للمتغير العشوائي المستمر X.

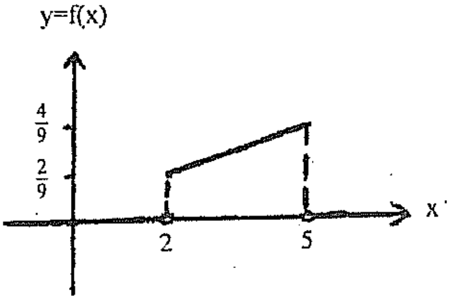
3- أوجد $P(X < 3)$.

4- أوجد $P(3 \le X < 4)$.

الحل: 1- يمكن التعبير عن دالة الكثافة المفروضة كما يلي:

$$f(x) = \begin{cases} \frac{2}{27}(1+x) & , \quad 2 \le x \le 5 \\ 0 & , \quad x < 2 \quad \text{أو} \quad x > 5 \end{cases}$$

حيث يمكن تمثيلها بيانياً كما في الشكل (6-3).



الشكل (6-3)

من تعريف دالة الكثافة $f(x)$ نلاحظ بسهولة أنّ $f(x) \ge 0$ أياً كان x من قيم المتغير العشوائي X. كما نلاحظ أنّ:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{2} f(x)dx + \int_{2}^{5} f(x)dx + \int_{5}^{\infty} f(x)dx$$

$$= 0 + \frac{2}{27}\int_{2}^{5}(1+x)dx + 0$$

$$=\frac{2}{27}\left[x+\frac{x^2}{2}\right]_2^5=\frac{2}{27}\left(\frac{27}{2}\right)=1$$

وبالتالي فإنّ الدالة $f(x)$ السابقة هي دالة كثافة احتمالية.

2- لإيجاد دالة التوزيع المجمعة للمتغير العشوائي المستمر X الذي دالة كثافته $f(x)$ نجد أنّ:

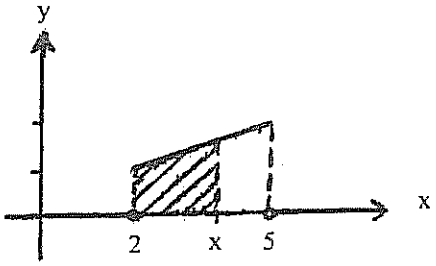
$$F(x)=P(X\le x)=\int_{-\infty}^{x}f(x)dx$$

$$=\int_{-\infty}^{2}f(x)dx+\int_{2}^{x}f(x)dx$$

$$=0+\frac{2}{27}\int_{2}^{x}(1+x)dx$$

$$=\frac{2}{27}\left[x+\frac{x^2}{2}\right]_2^x=\frac{2}{27}\left(x+\frac{x^2}{2}-4\right)$$

حيث يمكن توضيح ذلك من خلال الشكل (3-7)



الشكل (3-7)

3- لحساب الاحتمال $P(X<3)$ نلاحظ أنّ:

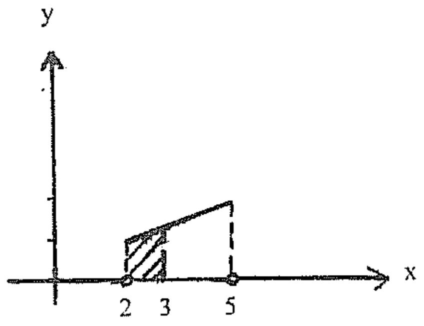
$$P(X<3)=\int_{-\infty}^{3}f(x)dx=\frac{2}{27}\int_{2}^{3}(1+x)dx=\frac{2}{27}\left[x+\frac{x^2}{2}\right]_2^3$$

$$=\frac{7}{27}=0.259$$

كما يمكننا حساب هذا الاحتمال عن طريق دالة التوزيع المجمعة كما يلي:

$$P(X<3)=P(X\le 3)=F(3)=\frac{2}{27}(3+\frac{9}{2}-4)=\frac{7}{27}=0.259$$

حيث يمكن توضيح ذلك من خلال الشكل (3-8)



الشكل (3-8)

4- لحساب الاحتمال $P(3 < X \leq 4)$ نلاحظ أن:

$$P(3 < X \leq 4) = \int_3^4 f(x)dx = \frac{2}{27}\int_3^4 (1+x)dx = \frac{2}{27}\left[x + \frac{x^2}{2}\right]_3^4$$

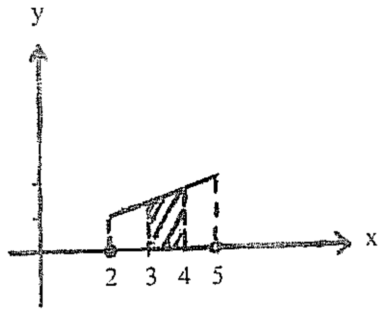
$$= (\frac{2}{27})(\frac{9}{2}) = \frac{9}{27} \cong 0.33$$

ويمكننا حساب الاحتمال السابق عن طريق دالة التوزيع المجمعة كما يلي:

$$P(3 < X \leq 4) = P(3 \leq X \leq 4) = F(4) - F(3)$$

$$= 0.593 - 0.259 \cong 0.33$$

حيث يمكن توضيح ذلك من خلال الشكل (3-9).



الشكل (3-9)

3-4 التوزيع المشترك لمتغيرين عشوائيين:

وجدنا أنه يمكن تحديد متغير عشوائي ما بمجرد معرفتنا بدالة الكثافة الاحتمالية أو دالة التوزيع المجمعة لهذا المتغير. ولكن هناك حالات إحصائية نضطر فيها إلى استخدام متغيرين عشوائيين أو أكثر حيث تكون هذه المتغيرات العشوائية جميعها معرفة على نفس فضاء التجربة وسنهتم بشكل خاص في الحالات التي تكون فيها هذه المتغيرات مستقلة عن بعضها البعض.

3.4.1 المتجه العشوائي المنقطع:

يُدعى المتجه العشوائي (X,Y) متجهاً عشوائياً منقطعاً، إذا كانت أياً من مركباته عبارة عن متغير عشوائي منقطع. أي أن المتجه (X,Y) يكون منقطعاً إذا كانت قيمهُ تشكل مجموعة منتهية أو قابلة للعد من الأزواج المرتبة من الأعداد الحقيقية.

3.4.2 التوزيع الاحتمالي المنقطع للمتجه العشوائي (X,Y):

إنّ الدالة المعرفة على مجموعة قيم المتجه العشوائي المنقطع (X,Y)، والتي تأخذ قيمها من مجموعة الأعداد الحقيقية R، والممثلة لاحتمال كل قيمة (x,y) من قيم (X,Y) تُدعى بدالة الكثافة الاحتمالية المشتركة للمتغيرين العشوائيين المنقطعين X و Y. يُرمز لهذه الدالة بالرمز $f(x,y)$ حيث تُعرف بالعلاقة:

$$f(x,y) = P(X = x, Y = y) \qquad (3.5)$$

إنّ هذه الدالة تعطى عادة على شكل جدول يُدعى جدول التوزيع الاحتمالي المنقطع للمتجه العشوائي (X,Y) والمبين كما يلي:

X \ Y	y_1 y_2 $\cdots$ y_m	المجموع
x_1 x_2 $\vdots$ x_n	$f(x_1,y_1)$ $f(x_1,y_2)\cdots\cdots f(x_1,y_m)$ $f(x_2,y_1)$ $f(x_2,y_2)\cdots\cdots f(x_2,y_m)$ .... $f(x_n,y_1)$ $f(x_n,y_2)\cdots\cdots f(x_n,y_m)$	$f(x_1)$ $f(x_2)$ $f(x_n)$
المجموع	$g(y_1)$ $g(y_2)$ ...... $g(y_m)$	1

حيث يمثل $f(x_i)$ مجموع كل الاحتمالات الواقعة في السطر المقابلة للقيمة x_i $(i=1,\cdots,n)$، ويمثل $g(y_j)$ مجموع كل الاحتمالات الواقعة في العمود المقابلة للقيمة y_j $(j=1,2,\cdots,m)$. تسمى الدوال $f(x)$ و $g(y)$ المعرفة بالعلاقتين التاليتين:

$$f(x)=\sum_y f(x,y) \quad , \quad g(y)=\sum_x f(x,y) \qquad \textbf{(3.6)}$$

بالدوال الهامشية (Marginal) لكل من X و Y على الترتيب. حيث تحقق دالة الكثافة الاحتمالية المشتركة (أو ما يسمى قانون التوزيع المشترك) الشرطين التاليين:

1- $f(x,y)\geq 0$

2- $\sum_x\sum_y f(x,y)=1$

ويكون المتجه العشوائي المنقطع (X,Y) محدداً تماماً إذا كان قانون التوزيع المشترك للمتغيرين العشوائيين المنقطعين X و Y معروفاً.

إنّ التمثيل البياني للمتجه العشوائي المنقطع (X,Y) نحصل عليه بإنشاء عمود في كل نقطة من المستوي بحيث يكون طول هذا العمود مساوياً الاحتمال المفروض بهذه النقطة. إنّ هذا التمثيل يعطي الخط البياني لدالة الكثافة الاحتمالية المشتركة للمتغيرين المنقطعين X و Y والتي عرفت بالعلاقة: $f(x,y)=P(X=x,Y=y)$

مثال (6): بفرض أنّ المتغيرين العشوائيين X و Y يأخذان القيم 1 و 2 و 3 وأنّ الاحتمالات التالية معطاة:

$P(X=1)=0.5$ $\qquad$ $P(X=2)=0.3$

$P(Y=1)=0.7$ $\qquad$ $P(Y=2)=0.2$

$P(X=1,Y=1)=0.35$ $\qquad$ $P(X=2,Y=2)=0.06$

$$P(X=3,Y=1)=0.2$$

أوجد جدول التوزيع الاحتمالي المشترك للمتغيرين العشوائيين X و Y.

الحل: اعتماداً على تعريف الدوال الهامشية لكل من المتغيرين المنقطعين X و Y لدينا:

$$P(X=1)+P(X=2)+P(X=3)=1$$

$$P(Y=1)+P(Y=2)+P(Y=3)=1$$

بالاستفادة من القيم المعطاة في المثال نجد أنّ:

$$P(X=3)=0.2\ ,\ P(Y=3)=0.1$$

وبما أنّ:

$$P(X=1)=P(X=1,Y=1)+P(X=1,Y=2)+P(X=1,Y=3)$$
$$P(X=2)=P(X=2,Y=1)+P(X=2,Y=2)+P(X=2,Y=3)$$
$$P(X=3)=P(X=3,Y=1)+P(X=3,Y=2)+P(X=3,Y=3)$$

وبالتعويض بالقيم المعطاة في العلاقات السابقة نجد أنّ:

$$\left.\begin{array}{l} 0.5=0.35+P(X=1,Y=2)+P(X=1,Y=3) \\ 0.3=P(X=2,Y=1)+0.06+P(X=2,Y=3) \\ 0.2=0.2+P(X=3,Y=2)+P(X=3,Y=3) \end{array}\right\} (1)$$

كذلك لدينا:

$$P(Y=1)=P(X=1,Y=1)+P(X=2,Y=1)+P(X=3,Y=1)$$
$$P(Y=2)=P(X=1,Y=2)+P(X=2,Y=2)+P(X=3,Y=2)$$
$$P(Y=3)=P(X=1,Y=3)+P(X=2,Y=3)+P(X=3,Y=3)$$

وبالتعويض بالقيم المعطاة في هذه العلاقات نجد أنّ:

$$\left.\begin{array}{l} 0.7=0.35+P(X=2,Y=1)+0.2 \\ 0.2=P(X=1,Y=2)+0.06+P(X=3,Y=2) \\ 0.1=P(X=1,Y=3)+P(X=2,Y=3)+P(X=3,Y=3) \end{array}\right\} (2)$$

بحل جمل المعادلات (1) و (2) نجد أنّ:

$$P(X=1,Y=2)=0.14 \qquad P(X=1,Y=3)=0.01$$
$$P(X=2,Y=1)=0.15 \qquad P(X=2,Y=3)=0.09$$
$$P(X=3,Y=2)=0 \qquad P(X=3,Y=3)=0$$

وهكذا فإنّ جدول التوزيع الاحتمالي المشترك للمتغيرين العشوائيين X و Y هو:

X \ Y	1	2	3	المجموع
1	0.35	0.14	0.01	0.50
2	0.15	0.06	0.09	0.30
3	0.20	0	0	0.20
المجموع	0.70	0.20	0.10	1

إنّ الجدول السابق يدعى أيضاً قانون التوزيع الاحتمالي المشترك لـ X و Y.

مثال (7): يحتوي جهاز الكتروني لا يعمل على تسع وصلات ثلاث منها بيضاء واثنتين سوداء وأربع حمراء. سحبنا عشوائياً وصلتين من الجهاز لفحصهما. ليكن X المتغير العشوائي الدال على عدد القطع البيضاء المسحوبة و Y المتغير العشوائي الدال على عدد القطع السوداء المسحوبة والمطلوب:

(1) عين جدول التوزيع الاحتمالي المشترك للمتغيرين العشوائيين X و Y.

(2) أوجد التوزيعات الهامشية لكل من المتغيرين المنقطعين X و Y.

الحل: (1) إنّ قانون التوزيع المشترك للمتغيرين العشوائيين X و Y يعطى بالشكل:

X \ Y	0	1	2
0	f(0,0)	f(0,1)	f(0,2)
1	f(1,0)	f(1,1)	f(1,2)
2	f(2,0)	f(2,1)	f(2,2)

حيث نجد أنّ:

$$f(0,0) = P(X=0, Y=0) = \frac{C_3^0 C_2^0 C_4^2}{C_9^2} = \frac{6}{36}$$

$$f(0,1) = P(X=0, Y=1) = \frac{C_3^0 C_2^1 C_4^1}{C_9^2} = \frac{8}{36}$$

$$\vdots$$

$$f(2,2) = P(X=2, Y=2) = 0$$

وبالتالي فإنّ جدول التوزيع الاحتمالي المشترك للمتغيرين X و Y هو:

X \ Y	0	1	2	المجموع
0	$6/36$	$8/38$	$1/36$	$15/36$
1	$12/36$	$6/36$	0	$18/36$
2	$3/36$	0	0	$3/36$
المجموع	$21/36$	$14/36$	$1/36$	1

(2) بفرض أنّ f(x) هي دالة الكثافة الهامشية لـ X و g(y) دالة الكثافة الهامشية لـ Y فإننا نجد من جدول التوزيع المشترك أنّ:

$$f(0)=P(X=0)=\sum_y P(X=0,Y=y)=P(X=0,Y=0)$$

$$+P(X=0,Y=1)+P(X=0,Y=2)=15/36$$

وأنّ:
$$f(1)=P(X=1)=18/36 \quad , \quad f(2)=P(X=2)=3/36$$

وهكذا فإنّ جدول التوزيع الاحتمالي المنقطع للمتغير X هو:

x	0	1	2
$f(x)=P(X=x)$	$15/36$	$18/36$	$3/36$

بنفس الطريقة نجد أنّ:

$$g(0)=P(Y=0)=\sum_x P(X=x,Y=0)=$$

$$P(X=0,Y=o)+P(X=1,Y=0)+P(X=2,Y=0)=21/36$$

وأنّ:
$$g(1)=14/36 \quad , \quad g(2)=1/36$$

وهكذا فإنّ جدول التوزيع الاحتمالي المنقطع للمتغير Y هو:

y	0	1	2
$g(y)=P(Y=y)$	$21/36$	$14/36$	$1/36$

3-4-3 المتجه العشوائي المستمر:

يدعى المتجه العشوائي(X,Y) متجهاً عشوائياً مستمراً إذا كانت كل مركبة من مركباته متغيراً عشوائياً مستمراً. أي يكون المتجه (X,Y) مستمراً إذا كانت قيمه تشكل مجموعة غير منتهية وغير قابلة للعد من الأزواج المرتبة من الأعداد الحقيقية.

3-4-4 دالة الكثافة الاحتمالية للمتجه العشوائي المستمر:

ليكن (X,Y) متجهاً عشوائياً مستمراً عندئذ تكون الدالة $f(x,y)$ المعرفة على مجموعة قيم هذا المتجه والتي تأخذ قيمها من مجموعة الأعداد الحقيقية R، دالة كثافة احتمالية للمتجه العشوائي المستمر (X,Y) إذا حققت الشرطين التاليين:

-1 $f(x,y)\geq 0$

-2 $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1$

تسمى الدالة $f(x,y)$ أيضاً دالة الكثافة الاحتمالية المشتركة للمتغيرين العشوائيين المستمرين X و Y.

من التعريف السابق نجد أنّ:

$$P(x_1 \le X \le x_2, y_1 \le Y \le y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f(x,y)dxdy \qquad (3.7)$$

وذلك أياً كانت الأعداد الحقيقية x_1, x_2, y_1, y_2 بحيث $x_1 \le x_2, y_1 \le y_2$. إنّ التمثيل البياني للمتجه العشوائي المستمر (X,Y) نحصل عليه برسم دالة الكثافة $f(x,y)$ بيانياً من أجل كل قيمة (x,y) في المستوي، وبالتالي فإنّ العلاقة (3.7) تمثل الحجم الواقع تحت منحني دالة الكثافة وفوق المنطقة المحدودة بالمستقيمات $x = x_1, x = x_2$ و $y = y_1, y = y_2$ في المستوي الديكارتي.

من تعريف دالة الكثافة $f(x,y)$ للمتجه المستمر (X,Y) نجد أنه يمكننا الحصول على دالتي الكثافة لكلٍ من المتغيرين العشوائيين المستمرين X و Y كما يلي:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad (3.8)$$

$$g(y) = \int_{-\infty}^{\infty} f(x,y)dx \qquad (3.9)$$

حيث تسمى الدالتان السابقتان $f(x)$ و $g(y)$ بدالتي الكثافة الهامشيتين للمتغيرين العشوائيين X و Y على الترتيب.

3-4-5 دالة التوزيع المجمعة (التراكمية) لمتجه عشوائي:

إذا كان X و Y متغيرين عشوائيين وكانت $f(x,y)$ دالة الكثافة المشتركة لهذين المتغيرين عندئذ تُعرف دالة التوزيع المجمعة للمتجه العشوائي (X,Y) والتي يُرمز لها بالرمز $F(x,y)$ بالعلاقة التالية:

$$F(x,y) = P[X \le x, Y \le y] \qquad (3.10)$$

حيث تمثل x و y أعداداً حقيقية من قيم المتغيرين X و Y على الترتيب. فإذا كان (X,Y) متجهاً عشوائياً منقطعاً فإنّ:

$$F(x_0,y_0) = P[X \le x_0, Y \le y_0] = \sum_{x\le x_0}\sum_{y \le y_0} f(x,y) = \sum_{x\le x_0}\sum_{y\le y_0} P[X = x, Y = y]$$

أمّا إذا كان (X,Y) متجهاً عشوائياً مستمراً فإنّ:

$$F(x_0,y_0)=P[X\le x_0,Y\le y_0]=p(X<x_0,Y<y_0)=\int_{-\infty}^{x_0}\int_{-\infty}^{y_0}f(x,y)dxdy$$

من العلاقة (3.10) نلاحظ أنّ دالة التوزيع المجمعة هي عبارة عن دالة في المتغيرين الحقيقيين x و y، ففي الحالة التي يكون فيها المتجه العشوائي (X,Y) مستمراً نلاحظ أنّ العلاقة الممثلة لدالة التوزيع المجمعة لهذا المتجه نحصل عليها من التكامل:

$$F(x,y)=P[X\le x,Y\le y]=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv \qquad \textbf{(3.11)}$$

مثال (8): ليكن X و Y متغيرين عشوائيين مستمرين دالة كثافتهما المشتركة معرفة بالعلاقة:

$$f(x,y)=\frac{x}{4}(1+3y^2) \quad ; \quad 0<x<2 \quad , \quad 0<y<1$$

والمطلوب:

1- أثبت أنّ الدالة $f(x,y)$ هي دالة كثافة احتمالية.

2- أوجد دالتي الكثافة الهامشيتين لكل من X و Y.

3- أوجد دالة التوزيع المجمعة للمتجه العشوائي المستمر (X,Y)

4- أحسب الاحتمال $P[X<1\ ,\ Y<0.5]$

الحل: 1- يمكن كتابة دالة الكثافة السابقة بالشكل:

$$f(x,y)=\begin{cases}\frac{3}{4}(1+3y^2) & ; \quad 0<x<2 \quad , \quad 0<y<1\\ 0 & ; \quad \text{عدا ذلك}\end{cases}$$

حيث نلاحظ وضوحاً أنّ $f(x,y)\ge 0$ أياً كان العددان الحقيقيان x و y. كما نلاحظ أنّ:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=1$$

لأنّ:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=\int_{0}^{2}\int_{0}^{1}\frac{x}{4}(1+3y^2)dxdy$$

$$=\frac{1}{4}\int_{0}^{2}x[\int_{0}^{1}(1+3y^2)dy]dx$$

$$=\frac{1}{4}\int_{0}^{2}x[y+y^3]_0^1dx=\frac{1}{4}\int_{0}^{2}2xdx=\frac{1}{2}[\frac{x^2}{2}]_0^2=1$$

2- لإيجاد دالتي الكثافة الهامشيتين لكلٍ من X و Y نجد أنّ:

$$f(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_{0}^{1}\frac{x}{4}(1+3y^2)dy=\frac{x}{4}[y+y^3]_0^1=\frac{x}{2}$$

$$g(y)=\int_{-\infty}^{\infty} f(x,y)dx=\int_{0}^{2}\frac{x}{4}(1+3y^2)dx=\frac{1+3y^2}{4}[\frac{x^2}{2}]_0^2=\frac{1+3y^2}{2}$$

وبالتالي فإنّ دالة الكثافة الهامشية لـ X هي:

$$f(x)=\begin{cases}\frac{x}{2} & , \quad 0<x<2\\ 0 & ; \quad \text{عدا ذلك}\end{cases}$$

ودالة الكثافة الهامشية لـ Y هي:

$$g(y)=\begin{cases}\frac{1+3y^2}{2} & , \quad 0<y<1\\ 0 & ; \quad \text{عدا ذلك}\end{cases}$$

3- لإيجاد دالة التوزيع المجمعة للمتجه (X,Y) نجد أنّ:

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv$$

$$=\int_{-\infty}^{x}\int_{-\infty}^{y}\frac{u}{4}u(1+3v^2)dudv$$

$$=\frac{1}{4}\int_{0}^{x}\int_{0}^{y}u(1+3v^2)dudv$$

$$=\frac{1}{4}\int_{0}^{x}u[v+v^3]_0^y du=\frac{1}{4}\int_{0}^{x}u[y+y^3]du$$

$$=\frac{1}{4}(y+y^3)[\frac{u^2}{2}]_0^x=\frac{1}{4}(y+y^3)[\frac{x^2}{2}]=\frac{1}{8}x^2(y+y^3)$$

4- لحساب الاحتمال $P[X<1\ ,\ Y<0.5]$ نجد أنّ:

$$P[X<1\ ,\ Y<0.5]=\int_{0}^{1}\int_{0}^{0.5}\frac{x}{4}(1+3y^2)dxdy$$

$$=\int_{0}^{1}\frac{x}{4}[y+y^3]_0^{0.5}dx=0.156[\frac{x^2}{2}]_0^1=0.078$$

حيث يمكن حساب الاحتمال السابق أيضاً عن طريق دالة التوزيع المجمعة كما يلي:

$$P[X<1\ ,\ Y<0.5]=F(1,0.5)=\frac{1}{8}(1)^2(0.5+(0.5)^3)=0.078$$

3-4-6 المتغيرات العشوائية المستقلة:

تؤدي المتغيرات العشوائية المستقلة دوراً كبيراً في الاستقراء الإحصائي، حيث نلاحظ أنّ معرفة دالة الكثافة الاحتمالية لكلٍ من المتغيرات العشوائية المستقلة يؤدي إلى معرفة دالة الكثافة المشتركة بينها. أمّا في حالة المتغيرات العشوائية غير المستقلة فنجد أنه لا يمكننا دوماً معرفة دالة الكثافة المشتركة لها بمجرد معرفة دالة الكثافة الاحتمالية لكلٍ منها.

تعريف (2):

إذا كانت $f(x,y)$ دالة الكثافة المشتركة للمتغيرين العشوائيين X و Y وكانت $f(x)$ و $g(y)$ دالتي الكثافة الهامشيتين لكلٍ من X و Y على الترتيب، عندئذٍ نقول عن المتغيرين العشوائيين X و Y أنهما مستقلان إذا وفقط إذا كان:

$$f(x,y) = f(x)g(y) \qquad (3.12)$$

يمكن تعميم التعريف السابق ليشمل n متغيراً عشوائياً. فإذا كانت $f(x_1, x_2, ..., x_n)$ دالة الكثافة المشتركة للمتغيرات العشوائية $X_1, X_2, ..., X_n$ وكانت $f_n(x_n), \cdots, f_1(x_1)$ دوال الكثافة الهامشية لكلٍ من $X_n, X_2, ..., X_1$ على الترتيب، عندئذٍ نقول عن المتغيرات العشوائية $X_n, ..., X_2, X_1$ أنها مستقلة إذا وفقط إذا كان:

$$f(x_1, x_2, ..., x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n) \qquad (3.13)$$

نتيجة (1):

إذا كان X و Y متغيرين عشوائيين كثافتهما المشتركة $f(x,y)$ ودالة توزيعهما المجمعة $F(x,y)$ فإنّ الشرط اللازم والكافي لكي يكون X و Y مستقلين هو:

$$F(x,y) = F(x)F(y) \qquad (3.14)$$

حيث تمثل $F(x)$ و $F(y)$ دوال التوزيع المجمعة لكل من المتغيرين X و Y على الترتيب. وفي الحالة التي تكون فيها $F(x_1, x_2, ..., x_n)$ دالة التوزيع المجمعة للمتجه العشوائي $(X_1, X_2, ..., X_n)$ الذي دالة كثافته $f(x_1, x_2, ..., x_n)$، فإنّ الشرط اللازم والكافي لكي تكون المتغيرات العشوائية $X_n, ..., X_2, X_1$ مستقلة هو:

$$F(x_1, x_2, ..., x_n) = F_1(x_1)F_2(x_2)\cdots F_n(x_n) \qquad (3.15)$$

حيث تمثل $F_n(x_n)\cdots, F_2(x_2), F_1(x_1)$ دوال التوزيع المجمعة لكل من المتغيرات العشوائية $X_n, ..., X_2, X_1$ على الترتيب.

مثال (9): بيّن فيما إذا كان المتغيران العشوائيان X و Y المعرفان في المثالين (7) و (8) مستقلين أم لا؟

الحل: بالعودة إلى المثالين المذكورين نلاحظ أن المتغيرين العشوائيين X و Y المعرفين في المثال (7) غير مستقلين بينما هما مستقلان في المثال (8) وذلك لتحقق الشرط $f(x,y)=f(x)g(y)$ في المثال (8) وعدم تحققه في المثال (7).

مثال (10): ليكن لدينا المتجه العشوائي المنقطع (X و Y) والذي قانون توزيعه الاحتمالي معرف بالشكل:

X \ Y	1	2	3
1	$1/9$	$1/9$	$1/9$
2	$1/9$	$1/9$	$1/9$
3	$1/9$	$1/9$	$1/9$

والمطلوب:

1- أوجد قانون التوزيع الاحتمالي لكلٍ من المتغيرين X و Y

2- أثبت أنّ X و Y متغيران عشوائيان مستقلان.

الحل: 1- من جدول التوزيع الاحتمالي المشترك نستنتج أن جدولي التوزيع الاحتماليين لكل من المتغيرين المنقطعين X و Y هما:

x	1	2	3
f(x)	$1/3$	$1/3$	$1/3$

y	1	2	3
g(y)	$1/3$	$1/3$	$1/3$

2- من قوانين التوزيع الاحتمالية للمتجه (X و Y) و كلٍ من المتغيرين X و Y نجد أنّ:

$f(x,y)=f(x)g(y)$ أياً كان $x,y=1,2,3$ أي أنّ المتغيرين X و Y مستقلان.

3-5 القيم المميزة للمتغيرات العشوائية:

نعلم أنّ دالة الكثافة الاحتمالية لمتغير عشوائي، تعطي وصفاً تاماً عن هذا المتغير ولكننا نحتاج في بعض الأحيان إلى ثوابت عددية أو وسطاء إحصائية تساعد في هذا الوصف للمتغير العشوائي، حيث أنّ معرفة القيم العددية لهذه الوسطاء الإحصائية تُمكن الباحث من الوصول إلى رؤية واضحة لطبيعة المتغيرات العشوائية.

في هذه الفقرة سندرس خمسة وسطاء إحصائية هي التوقع الرياضي لمتغير عشوائي أو شعاع عشوائي والتباين والانحراف المعياري لمتغير عشوائي، والتغاير ومعامل الارتباط بين متغيرين عشوائيين أو أكثر.

3-5-1 التوقع الرياضي لمتغير عشوائي (Mathematical expection of a random variable):

يعتبر التوقع الرياضي من أهم الوسطاء الإحصائية التي تصف متغيراً عشوائياً، وذلك لأنه يعطي الأساس المنطقي لمعظم طرائق الاستقراء الإحصائي ولأنه يدخل في تعريف معظم الوسطاء الإحصائية الأخرى.

تعريف (3):

إذا كان X متغيراً عشوائياً، فإنّ قيم هذا المتغير تتوزع عشوائياً حيث تميل غالباً إلى التراكم أو التجمع حول قيمة معينة تعبر عن تمركز هذا التوزيع، هذه القيمة هي القيمة المتوقعة (Expected value) أو التوقع الرياضي للمتغير العشوائي X حيث نرمز لها بالرمز $E(X)$.

وبصورة أخرى يمكننا تعريف التوقع الرياضي لمتغير عشوائي X على أنه المعدل النظري على المدى الطويل لقيمة المتغير X.

فإذا كان X متغيراً عشوائياً منقطعاً دالة كثافته الاحتمالية $f(x) = P(X = x)$ فإنّ التوقع الرياضي لهذا المتغير يكون:

$$\begin{aligned} E(X) &= \sum_x x f(x) \\ &= \sum_x x p(X = x) \end{aligned} \qquad (3.16)$$

وإذا كان X متغيراً عشوائياً مستمراً دالة كثافته الاحتمالية $f(x)$ فإنّ التوقع الرياضي للمتغير X في هذه الحالة يكون:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \qquad (3.17)$$

مثال (11): عائلة عندها أربعة أولاد. أكتب فضاء التجربة S. ثم بفرض X المتغير العشوائي الدال على عدد الإناث، أوجد جدول التوزيع الاحتمالي للمتغير العشوائي X. ثم أحسب التوقع الرياضي لهذا المتغير.

الحل: إذا رمزنا بالرمز B للذكر والرمز G للأنثى فإنّ فضاء التجربة في هذا المثال يكون:
S={BBBB,BGGG,GBBG,GBGG,GGBG,GGGB,BBGG,GBGB,BGBG,BGGB,GGBB,BBBG,BGBB,GGGG}.

ومنه فإن جدول التوزيع الاحتمالي المنقطع للمتغير العشوائي X هو:

x	0	1	2	3	4
$f(x) = P(X = x)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

حيث نجد أنّ التوقع الرياضي للمتغير العشوائي X هو:

$$E(X) = \sum_{X=0}^{4} xf(x) = (0)(1/16) + (1)(4/16) + (2)(6/16) + (3)(4/16) + (4)(1/16) = 32/16 = 2$$

وهذا يعني أننا نتوقع أن نجد بنتين في كل عائلة عندها أربعة أولاد.

مثال (12): جهاز ألكتروني يحوي تسعة ترانزستورات، ثلاثة منها لا تعمل، اخترنا بصورة عشوائية ثلاثة ترانزستورات، ونزعناها من الجهاز ثم فحصناها. بفرض أن X هو المتغير العشوائي الدال على عدد ما وجدناه عاطلاً منها. أوجد جدول التوزيع الاحتمالي للمتحول العشوائي X. ثم أحسب التوقع الرياضي لهذا المتغير.

الحل: إذا رمزنا بالرمز D للترانزستور التالف والرمز N للترانزستور السليم فإنّ فضاء التجربة في هذا المثال يكون:

S={DDD,DDN,DND,DNN,NDD,NDN,NND,NNN}

حيث نلاحظ أنّ المتغير العشوائي X يأخذ القيم 0، 1، 2، 3. ومنه نجد أنّ جدول التوزيع الاحتمالي المنقطع للمتغير العشوائي X هو:

x	0	1	2	3
$f(x) = P(X = x)$	$20/84$	$45/84$	$18/84$	$1/84$

وبالتالي فإنّ التوقع الرياضي لهذا المتغير يكون: $E(X)=\sum_{X=0}^{3} xf(x)=1$ أي أننـــا نتوقــع أن نجد ترانزستور واحد تالف ضمن الترانزستورات الثلاثة المختارة.

مثال (13): بالعودة إلى المثال (5) حيث كان لدينا X متغيراً عشوائياً مستمراً دالة كثافته:

$$f(x)=\tfrac{2}{27}(1+x) \quad , \quad 2\le x\le 5$$

ولنحسب التوقع الرياضي لهذا المتغير فنجد أنّ:

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx=\int_{-\infty}^{2} xf(x)dx+\int_{2}^{5} xf(x)dx+\int_{5}^{\infty} xf(x)dx$$

$$=\int_{2}^{5} x[\tfrac{2}{27}(1+x)]dx=\tfrac{2}{27}\int_{2}^{5}(x+x^2)dx=3.67$$

من التعريف السابق للتوقع الرياضي يُلاحظ بأنّ القيمة المتوقعة تمثل حالة عامة لما يسمى في الفيزياء بالعزم والذي يُعرف بحاصل ضرب القوة في ذراعها أو لما يسمى في مضمار العلوم الطبيعية بالزخم والذي يعرف بحاصل ضرب الكتلة في المسافة. وما يهمنا هنا هــو المفهــوم الإحصائي للقيمة المتوقعة والذي يعني كمية الاحتمال الكلية للمتغير العشوائي X في تجربــة معينة.

ملاحظة ③: يمكن تعميم مفهوم التوقع الرياضي لمتغير عشوائي منقطع أو مستمر ليشمل حالة حساب القيمة المتوقعة لدالة في هذا المتغير العشوائي. فإذا كان X متغيراً عشوائياً منقطعاً دالة كثافته الاحتمالية $f(x)$، وكانت $u(x)$ دالة عددية في X فإنّ التوقع الرياضي للدالـــة $u(X)$ يعطى بالعلاقة:

$$E[u(X)]=\sum_{x} u(x)f(x) \qquad (3.18)$$

وإذا كان X متغيراً عشوائياً مستمراً دالة كثافته الاحتمالية $f(x)$، وكانت $u(x)$ دالة عددية في X مستمرة على مجموعة الأعداد الحقيقية أو علـــى جـــزء منهـــا فـــإنّ التوقـــع الرياضـــي للدالة $u(X)$ يكون:

$$E[u(X)]=\int_{-\infty}^{\infty} u(x)f(x)dx \qquad (3.19)$$

في الحقيقة إنّ أي دالة عددية $u(x)$ بمتغير عشوائي X هي عبارة عن متغير عشوائي آخر هو $Y = u(X)$، فإذا كانت $g(y)$ هي دالة الكثافة الاحتمالية لهذا المتغير الجديد وكانت $F(y)$ دالة التوزيع المجمعة له، فإنّ القيمة المتوقعة للدالة $u(X)$ هي نفسها القيمة المتوقعة للمتغير العشوائي $Y = u(X)$. حيث نجد في الحالة التي يكون فيها المتغير العشوائي $Y = u(X)$ منقطعاً أنّ القيمة المتوقعة لـ Y هي:

$$E(Y) = \sum_{y} y g(y)$$

ودالة التوزيع المجمعة هي: $F(y) = P(Y \leq y) = \sum_{Y \leq y} g(y)$

أمّا في الحالة التي يكون فيها المتغير Y مستمراً فإنّ القيمة المتوقعة لـ Y تكون:

$$E(Y) = \int_{-\infty}^{\infty} y g(y) dy$$

كما تكون دالة التوزيع المجمعة للمتغير Y بالشكل:

$$F(y) = \int_{-\infty}^{y} g(y) dy$$

وبحالة خاصة عندما تكون $u(X) = X$ فإنّ القيمة المتوقعة للدالة $u(X)$ تسمى بالقيمة المتوقعة للمتغير العشوائي X.

3-5-2. التوقع الرياضي لمتجه عشوائي:

يمكن توسيع مفهوم وتعريف القيمة المتوقعة ليشمل عدداً من المتغيرات العشوائية مثل $X_1, X_2, ..., X_n$. فإذا كانت $f(x_1, x_2, ..., x_n)$ دالة الكثافة الاحتمالية للمتجه العشوائي $(X_1, X_2, ..., X_n)$ وكانت $u(x_1, x_2, ..., x_n)$ دالة عددية في المتجه $(X_1, X_2, ..., X_n)$ أي دالة عددية في المتغيرات العشوائية $X_1, X_2, ..., X_n$، فإنّ التوقع الرياضي للدالة $u(X_1, X_2, ..., X_n)$ يُعرف بإحدى العلاقتين التاليتين وذلك حسب كون المتجه العشوائي $(X_1, X_2, ..., X_n)$ منقطع أو مستمر على الترتيب:

$$E[u(X_1, X_2, ..., X_n)] = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} u(x_1, x_2, ..., x_n) f(x_1, x_2, ..., x_n)$$

$$E[u(X_1, X_2, ..., X_n)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, x_2, ..., x_n) f(x_1, x_2, ..., x_n) dx_1 \ldots dx_n$$

وبشكل خاص من أجل متغيرين عشوائيين X,Y نجد أنه إذا كانـت $f(x,y)$ دالـة الكثافـة الاحتمالية للمتجه (X,Y) وكانت $u(x,y)$ دالة عددية في X و Y فـإنّ التوقـع الرياضـي للدالة $u(X,Y)$ يكون:

$$E[u(X,Y)] = \sum_x \sum_y u(x,y)f(x,y) \qquad \textbf{(3.20)}$$

عندما يكون المتجه (X,Y) منقطعاً. وإذا كان المتجه (X,Y) مستمراً والدالة $u(x,y)$ مستمرة فإنّ التوقع الرياضي للدالة $u(X,Y)$ يكون:

$$E[u(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x,y)f(x,y)dxdy \qquad \textbf{(3.21)}$$

وفيما يلي نذكر بعض خواص التوقع الرياضي الهامة.

3-5-3 خواص التوقع الرياضي (أو القيمة المتوقعة):

1- القيمة المتوقعة لأي كمية ثابتة تساوي نفس تلك الكمية. أي أنّ:

$$E(k) = k \qquad \textbf{(3.22)}$$

حيث k يمثل عدداً ثابتاً. وذلك لأنّ:

$$E(k) = \begin{cases} \sum_x kf(x) = k\sum_x f(x) = k(1) = k \\ \int_{-\infty}^{\infty} kf(x) = k\int_{-\infty}^{\infty} f(x)dx = k(1) = k \end{cases}$$

2- إذا كانت $u(x)$ دالة في المتغير العشوائي المنقطع أو المستمر X فإنّ:

$$E[ku(X)] = kE[u(X)] \qquad \textbf{(3.23)}$$

حيث k يمثل عدداً ثابتاً.

أي أنّ القيمة المتوقعة لحاصل ضرب عدد ثابت في دالة عددية لمتغير عشوائي يساوي حاصل ضرب العدد الثابت في القيمة المتوقعة لتلك الدالة. ولإثبات ذلك نلاحظ في حالة كون المتغير العشوائي X منقطعاً و $f(x)$ دالة كثافته الاحتمالية أنّ:

$$E[ku(X)] = \sum_x ku(x)f(x)$$
$$= k\sum_x u(x)f(x) = kE[u(X)]$$

أمّا في حالة كون المتغير العشوائي X مستمراً و $f(x)$ دالة كثافته الاحتمالية والدالـة $u(x)$ مستمرة فنجد أنّ:

$$E[ku(X)] = \int_{-\infty}^{\infty} ku(x)f(x)dx = k\int_{-\infty}^{\infty} u(x)f(x)dx = kE[u(X)]$$

3- إذا كانت $u_1(x_1), u_2(x_2), \ldots, u_n(x_n)$ دوال عددية في المتغيرات العشوائية (المنقطعة أو المستمرة) $X_1, X_2, \ldots X_n$ على الترتيب فإنّ:

$$E(k_1u_1 + k_2u_2 + \ldots + k_nu_n) = k_1E(u_1) + k_2E(u_2) + \ldots k_nE(u_n) \qquad \textbf{(3.24)}$$

حيث تمثل $k_1, k_2, \ldots, k_n$ أعداداً ثابتة. أي أنّ القيمة المتوقعة هي عملية خطية (Linear Operator) . وبشكل خاص إذا كانت: $k_1 = k_2 = \ldots = k_n = 1$ فإنّ:

$$E(u_1 + u_2 + \ldots + u_n) = E(u_1) + E(u_2) + \ldots E(u_n) \qquad \textbf{(3.25)}$$

أي أنّ القيمة المتوقعة لمجموع عدد من الدوال يساوي مجموع القيم المتوقعة لتلك الدوال. من الخواص السابقة وفي الحالة الخاصة التي يكون فيها $u(X) = X$ و $u_1(X_1) = X_1, \ldots, u_n(X_n) = X_n$ نستنتج أنه:

$$E(kX) = kE(X) \qquad \textbf{(3.26)}$$

حيث k يمثل عدداً ثابتاً.

$$E(k_1X_1 + k_2X_2 + \ldots + k_nX_n) = k_1E(X_1) + k_2E(X_2) + \ldots + k_nE(X_n) \qquad \textbf{(3.27)}$$

حيث تمثل $k_1, k_2, \ldots, k_n$ أعداداً ثابتة. وبشكل خاص يكون:

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n) \qquad \textbf{(3.28)}$$

مثال **(14)**: إذا كانت الدالة:

$$f(x) = \begin{cases} \frac{1}{2}x \;, & x = 1,3,5,7 \\ 0 \;, & \text{عدا ذلك} \end{cases}$$

هي دالة الكثافة الاحتمالية للمتغير العشوائي X، فأحسب كُلاً مما يلي:

أ- $E(X)$

ب- $E(X^3)$

جـ- $E(2X + 3X^3)$

الحل: نلاحظ من تعريف الدالة $f(x)$ بأنّ المتغير العشوائي X هو متغير منقطع، حيث نجد أنّ:

أ- $$E(X) = \sum_x xf(x) = (1)(\frac{1}{2}) + (3)(\frac{3}{2}) + (5)(\frac{5}{2}) + (7)(\frac{7}{2}) = \frac{84}{2} = 42$$

ب- $$E(X^3)=\sum_x x^3f(x)=(1^3)(\frac{1}{2})+(3^3)(\frac{3}{2})+(5^3)(\frac{5}{2})+(7^3)(\frac{7}{2})$$
$$=\frac{3108}{2}=1554$$

جـ- بفرض $u(x)=2x+3x^3$ دالة عددية في المتغير العشوائي X فإنّ:

$$E[u(X)]=E[2X+3X^3]=\sum_x(2x+3x^3)f(x)$$
$$(5)(\frac{1}{2})+(87)(\frac{3}{2})+(385)(\frac{5}{2})+(1043)(\frac{7}{2})$$
$$=\frac{9492}{2}=4746$$

ويمكن حساب التوقع الرياضي للدالة $u(X)$ بطريقة أخرى وذلك بالاستفادة من خواص التوقع حيث نجد أنّ:

$$E[u(X)]=E[2X+3X^3]=2E(X)+3E(X^3)$$
$$=2(42)+3(1554)=4746$$

مثال (15): إذا كانت الدالة:

$$f(x)=\begin{cases}3(1-\frac{1}{2}x)\ , & 0<x<1\\ 0 & ,\quad \text{عدا ذلك}\end{cases}$$

هي دالة الكثافة الاحتمالية للمتغير العشوائي X، فأحسب القيمة المتوقعة للدالة العددية $u(X)$ في كلٍ من الحالات الآتية:

أ- $u(X)=X$

ب- $u(X)=X^2$

جـ- $u(X)=6X+3X^2$

الحل: نلاحظ من تعريف الدالة $f(x)$ بأنّ المتغير العشوائي X هو متغير مستمر، حيث نجد أنّ:

أ- $$E[u(X)]=E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_0^1 3x(1-\frac{1}{2}x)dx$$
$$=3\int_0^1 xdx-\frac{3}{2}\int_0^1 x^2dx=\frac{3}{2}-\frac{1}{2}=1$$

ب- $$E[u(X)] = E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 3x^2(1-\frac{1}{2}x)dx$$

$$= 3\int_0^1 x^2dx - \frac{3}{2}\int_0^1 x^3dx = 1 - \frac{3}{8} = \frac{5}{8}$$

جـ- $$E[u(X)] = E[6X + 3X^2] = \int_0^1 3(6x + 3x^2)(1-\frac{1}{2}x)dx$$

$$= 3\int_0^1 (6x - \frac{3}{2}x^3)dx = 3\left[\frac{6x^2}{2} - \frac{3}{2}\frac{x^4}{4}\right]_0^1 = 9(1-\frac{1}{8}) = \frac{63}{8} = 7.875$$

ويمكن حساب التوقع الرياضي الأخير بطريقة أخرى وذلك باستخدام الخاصية الخطية للقيمة المتوقعة كما يلي:

$$E[u(X)] = E[6X + 3X^2] = 6E(X) + 3E(X^2) = 6(1) + (3)(\frac{5}{8}) = \frac{63}{8} = 7.875$$

مثال (16): أوجد التوقع الرياضي للدالة $u(X,Y) = XY$ علماً بأنّ قانون التوزيع المشترك للمتغيرين العشوائيين المنقطعين X و Y مُعرّف بالجدول التالي:

X \ Y	1	2	3
1	0.35	0.14	0.01
2	0.15	0.06	0.09
3	0.20	0	0

ثم أوجد التوقع الرياضي للمتغير العشوائي X.

الحل: بما أن المتجه العشوائي (X,Y) منقطع فإنّ:

$$E[u(X,Y)] = E(XY) = \sum_x \sum_y xy \quad f(x,y)$$

$$= (1)(1)(0.35) + (1)(2)(0.14) + (1)(3)(0.01) +$$
$$(2)(1)(0.15) + (2)(2)(0.06) + (2)(3)(0.09) +$$
$$(3)(1)(0.20) + (3)(2)(0) + (3)(3)(0)$$
$$= 0.35 + 0.28 + 0.03 + 0.30 + 0.24 + 0.36 + 0.60 = 2.16$$

لنوجد الآن التوقع الرياضي للمتغير X أي للدالة $u(X,Y) = X$ فنجد أنّ:

$$E[u(X,Y)] = E(X) = \sum_x \sum_y x \quad f(x,y)$$

$$= 1(0.35 + 0.14 + 0.01) + 2(0.15 + 0.06 + 0.09)$$

$$+ 3(0.20) = 0.35 + 0.14 + 0.01 + 0.30 + 0.12 + 0.18 + 0.60 = 1.7$$

حيث نلاحظ أنه يمكن حساب التوقع الرياضي السابق للمتغير X بطريقة أخرى وذلك من خلال جدول التوزيع الاحتمالي المنقطع لهذا المتغير كما يلي:

من قانون التوزيع المشترك للمتغيرين X وY المبين أعلاه نجد أنّ جدول التوزيع الاحتمالي المنقطع للمتغير X (أو قانون توزيع X) هو:

x	1	2	3
f(x)	0.5	0.3	0.2

ومنه نجد أنّ: $E(X) = \sum_{x} xf(x) = (1)(0.5) + (2)(0.3) + (3)(0.2) = 1.7$

وهو نفس الجواب الذي حصلنا عليه بالطريقة الأولى.

مثال (17): إذا كانت الدالة:

$$f(x,y,z) = \begin{cases} (x+y)z\,;\; 0 < x < 1\,,\; 0 < y < z < 1 \\ 0 \quad ; \quad \text{عدا ذلك} \end{cases}$$

هي دالة الكثافة الاحتمالية للمتجه العشوائي (X,Y,Z) ، فأحسب القيمة المتوقعة لكلاً مما يلي:

أ- $U(X,Y,Z) = (X^2 + YZ)$

ب- $U(X,Y,Z) = X + Y + Z$

الحل: بما أن المتجه العشوائي (X,Y,Z) مستمر فإنّ:

أ -

$$E[u(X,Y,Z)] = E(X^2 + YZ)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x^2 + yz)f(x,y,z)dxdydz$$

$$= \int_0^1\int_0^z\int_0^1 (x^2 + yz)(x+y)zdxdydz$$

$$= \int_0^1\int_0^z\int_0^1 (x^3z + x^2yz + yz^2x + y^2z^2)dxdydz$$

$$=\int_0^1\int_0^z(\frac{z}{4}+\frac{yz}{3}+\frac{yz^2}{2}+y^2z^2)dydz$$

$$=\int_0^1(\frac{z^2}{4}+\frac{z^3}{6}+\frac{z^4}{4}+\frac{z^5}{3})dz$$

$$=\left[\frac{z^3}{12}+\frac{z^4}{24}+\frac{z^5}{20}+\frac{z^6}{18}\right]_0^1=\frac{1}{12}+\frac{1}{24}+\frac{1}{20}+\frac{1}{18}=0.231$$

ب –

$$E[u(X,Y,Z)]=E(X+Y+Z)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y+z)f(x,y,z)dxdydz$$

$$=\int_0^1\int_0^z\int_0^1(x+y+z)(x+y)zdxdydz$$

$$=\int_0^1\int_0^z\int_0^1(x^2z+2xyz+y^2z+xz^2+yz^2)dxdydz$$

$$=\int_0^1\int_0^z(\frac{z}{3}+yz+y^2z+\frac{z^2}{2}+yz^2)dydz$$

$$=\int_0^1(\frac{z^2}{3}+\frac{z^3}{2}+\frac{z^4}{3}+\frac{z^3}{2}+\frac{z^4}{2})dz$$

$$=\int_0^1(\frac{z^2}{3}+z^3+\frac{5z^4}{6})dz$$

$$=\left[\frac{z^3}{9}+\frac{z^4}{4}+\frac{z^5}{6}\right]_0^1=\frac{1}{9}+\frac{1}{4}+\frac{1}{6}=0.528$$

النظرية التالية توضح العلاقة بين القيمة المتوقعة والاستقلالية للمتغيرات العشوائية.

نظرية ⓫:

إذا كانت $f(x_1)$ و $f(x_2)$ دالتي الكثافة الهامشيتين لكل من المتغيرين العشـوائيين X_1وX_2 على الترتيب وإذا كان المتغيران X_1وX_2 مستقلين فإنّ القيمة المتوقعة لحاصل ضرب الدالتين العدديتين $g(X_1)$ و $h(X_2)$ يساوي حاصل ضرب قيمتهما المتوقعتين. أي أنّ:

$$E[g(X_1)h(X_2)]=E[g(X_1)]E[h(X_2)] \qquad \textbf{(3.29)}$$

الاثبات:

لنفترض أنّ المتغيرين العشوائيين X_1وX_2 مستقلين عندئذ نجد أنّ:

$$f(x_1,x_2)=f(x_1)f(x_2)$$

حيث تمثل $f(x_1,x_2)$ دالة الكثافة المشتركة للمتغيرين X وY وبفرض أنّ المتغيرين X_1وX_2 مستمرين ومن تعريف القيمة المتوقعة نجد أنّ:

$$E[g(X_1)h(X_2)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x_1)h(x_2)f(x_1,x_2)dx_1dx_2$$

وبما أنّ x_1,x_2 مستقلان فإنّ:

$$E[g(X_1)h(X_2)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x_1)h(x_2)f(x_1)f(x_2)dx_1dx_2$$

$$=\left(\int_{-\infty}^{\infty}g(x_1)f(x_1)dx_1\right)\left(\int_{-\infty}^{\infty}h(x_2)f(x_2)dx_2\right)$$

$$=E[g(X_1)]E[h(X_2)]$$

وبنفس الأسلوب يمكن إثبات النظرية إذا افترضنا أنّ المتغيرين المستقلين X_1وX_2 غير مستمرين أو منقطعين.

نتيجة (2):

إذا كانت المتغيرات العشوائية $X_1,X_2,...,X_n$ مستقلة فإنّ القيمة المتوقعة لحاصل ضرب الدوال العددية $g_1(X_1),\ldots,g_n(X_n)$ تساوي حاصل ضرب القيم المتوقعة لكل من هذه الدوال. أي أنّ:

$$E[g_1(X_1).g_2(X_2)\ldots g_n(X_n)]=E[g_1(X_1)].E[g_2(X_2)]\ldots E[g_n(X_n)] \quad \textbf{(3.30)}$$

نتيجة (3):

إذا كانت الدوال العددية $g_1(X_1),\ldots,g_n(X_n)$ معطاة بالشكل:

$$g_1(X_1)=X_1, g_2(X_2)=X_2,\ldots, g_n(X_n)=X_n$$

عندئذ نجد أنه عندما تكون المتغيرات العشوائية $X_1,...,X_n$ مستقلة فإنّ:

$$E(X_1.X_2,...,X_n)=E(X_1).E(X_2)\ldots E(X_n) \qquad \textbf{(3.31)}$$

مثال (18): بالعودة إلى المثال (8) حيث كان لدينا X وY متغيرين عشوائيين دالة كثافتهما المشتركة هي:

$$f(x,y)=\frac{x}{4}(1+3y^2) \quad ; \quad 0<x<2 \quad , \quad 0<y<1$$

وكانت دوال الكثافة الهامشية لكل من المتغيرين X وY هي:

$$f(x)=\frac{x}{2} \quad ; \quad 0<x<2$$

$$g(y)=\frac{1+3y^2}{2} \quad ; \quad 0<y<1$$

حيث وجدنا في المثال (9) أنّ المتغيرين العشوائيين X وY مستقلان لأنّ:

$$f(x,y)=f(x)g(y)$$

لنحسب الآن التوقع الرياضي لكل من $X,\ Y,\ XY$. حيث نجد أنّ:

$$E(XY)=\int_0^2\int_0^1 xy\frac{x}{4}(1+3y^2)dxdy$$

$$=\int_0^2\frac{x^2}{4}\left(\int_0^1 y(1+3y^2)dy\right)dx=\frac{5}{6}$$

$$E(X)=\int_0^2\int_0^1 x\frac{x}{4}(1+3y^2)dxdy$$

$$=\int_0^2\frac{x^2}{4}\left(\int_0^1 (1+3y^2)dy\right)dx=\frac{4}{3}$$

$$E(Y)=\int_0^2\int_0^1 y\frac{x}{4}(1+3y^2)dxdy$$

$$=\int_0^2\frac{x}{4}\left(\int_0^1 y(1+3y^2)dy\right)dx=\frac{5}{6}$$

ويمكن حساب التوقع الرياضي لكل من X وY مباشرة عن طريق دالتي الكثافة الهامشيتين f(x) وg(y) حيث نجد أنّ:

$$E(X)=\int_0^2\frac{x^2}{2}dx=\left[\frac{x^3}{6}\right]_0^2=\frac{4}{3}$$

$$E(Y)=\int_0^1 y\frac{1+3y^2}{2}dy=\frac{5}{8}$$

ممّا سبق نجد أنّ: $E(XY)=\frac{5}{6}=E(X).E(Y)=\frac{4}{3}.\frac{5}{8}$ وذلك لأن X وY متغيران عشوائيان مستقلان.

3-5-4 القيمة المتوقعة والمتوسط الحسابي لمتغير عشوائي:

ليكن X متغيراً عشوائياً منقطعاً. فإنّ المتوسط الحسابي لقيم X (أو لـ X) هو:

$$\overline{X}=\frac{\sum_{i=1}^{n}x_i}{n}$$

وعندما يكون التوزيع لقيم X فئوياً نجد أنّ المتوسط الحسابي يساوي:

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{k}x_i f(x_i)$$

حيث تمثل $f_i=f(x_i)$ تكرار المركز x_i $(i=1,2,\ldots,k)$ وتمثل k عدد الفئات. ومنه يمكن كتابة المتوسط الحسابي بالشكل: $\overline{X}=\sum_{i=1}^{k}x_i\frac{f(x_i)}{n}$

حيث تمثل $\frac{f(x_i)}{n}$ التكرار النسبي للقيمة x_i $(i=1,2,\ldots,k)$ وهو نفسه قيمة دالة الكثافة الاحتمالية للمتغير العشوائي X عند النقطة x_i أي أنّ: $\frac{f(x_i)}{n}=P(X=x_i)$

ومنه فإنّ: $\overline{X}=\sum_{i=1}^{k}x_i P(X=x_i)=E(X)$

حيث نرمز عادة للمتوسط الحسابي لمتغير عشوائي X بالرمز μ_X وبالتالي فإنّ:

$$\mu_X=E(X) \qquad \textbf{(3.32)}$$

ممّا سبق نجد أنّ القيمة المتوقعة أو التوقع الرياضي لمتغير عشوائي منقطع هو ليس إلاّ المتوسط الحسابي لهذا المتغير.

إنّ العلاقة (3.32) تبقى محققة في كل من حالتي المتغير العشوائي المنقطع والمستمر.

مثال (19): يحتوي جهاز الكتروني على 36 وصلة كهربائية سُجلت أطوالها ضمن الجدول الفئوي التالي:

x_i الطول	$f(x_i)$ التكرار
20 cm	5
30 cm	7
40 cm	8
50 cm	6
60 cm	6
70 cm	4

بفرض X المتغير العشوائي الدال على طول الوصلة. أحسب المتوسط الحسابي للمتغيــر X وكذلك التوقع الرياضـي له. ماذا تلاحظ؟

الحل: من الجدول الفئوي نلاحظ أنّ المتوسط الحسابي للمتغير X (لقيم المتغير X) هو:

$$\overline{X}=\frac{\sum_{i=1}^{6}x_i f(x_i)}{\sum_{i=1}^{6}f(x_i)}=\frac{\sum_{i=1}^{6}x_i f(x_i)}{n}$$

$$=\frac{(20)(5)+(30)(7)+(40)(8)+(50)(6)+(60)(6)+(70)(4)}{36}=\frac{1570}{36}=\frac{785}{18}$$

من جهة أخرى نجد أنّ جدول التوزيع الاحتمالي المنقطع للمتغير العشوائي X هو:

x	20	30	40	50	60	70
$P(X=x)$	$5/36$	$7/36$	$8/36$	$6/36$	$6/36$	$4/36$

ومنه فإنّ التوقع الرياضي للمتغير X يكون:

$$E(X)=\sum_x xP(X=x)=(20)(5/36)+(30)(7/36)+$$

$$+(40)(8/36)+(50)(6/36)+(60)(6/36)+(70)(4/36)=\frac{1570}{36}=\frac{785}{18}$$

أي أنّ: $\overline{X}=E(X)$

3-5-5 تباين (تشتت) المتغير العشوائي (Variance of a random variable):

إنّ إيجاد التوقع الرياضي لمتغير عشوائي لا يفي بالغرض في كثير من الأحيان حيث غالباً ما نحتاج لمعرفة العلاقة بين قيم المتغير العشوائي والقيمة المتوقعة له.

تعريف (4):

ليكن X متغيراً عشوائياً توقعه الرياضي $E(X) = \mu_x = \mu$. إنّ تباين المتغير العشوائي X والذي نرمز له بالرمز Var(X) أو بالرمز σ_x^2 يعطى بالعلاقة:

$$Var(X) = \sigma_x^2 = E(X-\mu)^2 \qquad (3.33)$$

فإذا كان X متغيراً عشوائياً منفصلاً دالة كثافته الاحتمالية $f(x)$ فإنّ:

$$Var(X) = \sum_x (x-\mu)^2 f(x) \qquad (3.34)$$

أمّا إذا كان X متغيراً عشوائياً مستمراً دالة كثافته الاحتمالية $f(x)$ فإنّ:

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \qquad (3.35)$$

تعريف (5) الانحراف المعياري لمتغير عشوائي:

إذا كان X متغيراً عشوائياً تباينهُ $\sigma_x^2 = Var(X)$ فإننا ندعو الجذر التربيعي الموجب للتباين بالانحراف المعياري للمتغير العشوائي X. أي أنّ:

$$\sigma_x = \sqrt{Var(X)} \qquad (3.36)$$

مثال (20): مجموعة مكونة من 6 طلاب من كلية الهندسة الكهربائية والميكانيكية بجامعة دمشق بينهم اثنان من قسم الطاقة. اخترنا اثنين من المجموعة بصورة عشوائية. المتغير العشوائي X يدل على عدد طلاب قسم الطاقة في المجموعة المختارة. والمطلوب:

أ) أوجد جدول التوزيع الاحتمالي المنقطع للمتغير العشوائي X.

ب) أحسب التوقع الرياضي للمتغير العشوائي X.

جـ) أحسب تباين المتغير العشوائي X وانحرافه المعياري.

الحل: أ) إن جدول التوزيع الاحتمالي المنقطع للمتغير X هو:

x	0	1	2
f(x)	$6/15$	$8/15$	$1/15$

ب) التوقع الرياضي للمتغير العشوائي X أو المتوسط الحسابي له هو:

$$E(X) = \mu_x = \sum_x xf(x) = (0)(\frac{6}{15}) + (1)(\frac{8}{15}) + (2)(\frac{1}{15}) = \frac{2}{3}$$

جـ) التباين للمتغير العشوائي X هو:

$$Var(X) = \sigma_x^2 = \sum_x (x-\mu)^2 f(x)$$

حيث: $$\mu = \mu_x = \frac{2}{3}$$

ومنه: $$Var(X) = (0-\frac{2}{3})^2(\frac{6}{15}) + (1-\frac{2}{3})^2(\frac{8}{15}) + (2-\frac{2}{3})^2(\frac{1}{15})$$

$$= \frac{48}{135} = \frac{16}{45} = 0.36$$

وبالتالي فإنّ الانحراف المعياري للمتغير X هو:

$$\sigma_x = \sqrt{0.36} = 0.6$$

مثال (21): بالعودة إلى المثال (5) حيث كان لدينا X متغيراً عشوائياً مستمراً دالة كثافته:

$$f(x) = \frac{2}{27}(1+x) \quad , \quad 2 \le x \le 5$$

في المثال (13) وجدنا أن التوقع الرياضي لهذا المتغير هو: $\mu_x = E(X) = 3.67$

لنحسب الآن التباين للمتغير العشوائي X حيث نجد أنّ:

$$Var(X) = \sigma_x^2 = E(X-\mu)^2$$

$$= \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$$

$$= \int_{2}^{5}(x-3.67)^2 \frac{2}{27}(1+x)dx$$

$$= \frac{2}{27}\int_{2}^{5}(x^3-6.34x^2+6.13x+13.47)dx = 0.723$$

ومنه فإنّ الانحراف المعياري للمتغير X هو: $\sigma_x = \sqrt{0.723} = 0.850$

نتيجة (4):

من التعريف (4) لتباين متغير عشوائي X والاستفادة من خواص التوقع الرياضي نجد أنّ:

$$Var(X) = E(X-\mu)^2$$

$$= E(X^2 - 2\mu X + \mu^2)$$

$$= E(X^2) + E(-2\mu X) + E(\mu^2)$$

$$= E(X^2) - 2\mu E(X) + \mu^2$$

وبما أنّ: $\mu = E(X)$ فإنّ: $Var(X) = E(X^2) - 2\mu^2 + \mu^2$

ومنه:

$$Var(X) = E(X^2) - \mu^2 \qquad \textbf{(3.37)}$$

وبالتالي فإنه يمكننا حساب التباين لمتغير عشوائي X من العلاقة (3.37) أيضاً.

مثال **(22)**: ليكن X متغيراً عشوائياً منقطعاً له جدول التوزيع الاحتمالي التالي:

x	0	1	2	3
f(x)	$27/64$	$27/64$	$9/64$	$1/64$

أوجد التوقع الرياضي للمتغير X وكذلك التباين والإنحراف المعياري.

الحل: إنّ التوقع الرياضي للمتغير العشوائي المنقطع X هو:

$$\mu_x = E(X) = \sum_x x f(x) = (0)(27/64) + (1)(27/64) + (2)(9/64) + (3)(1/64) = 3/4 = 0.75$$

لنوجد الآن التوقع الرياضي لـ X^2 حيث نجد أنّ:

$$E(X^2) = \sum_x x^2 f(x) = (0)^2(27/64) + (1)^2(27/64) + (2)^2(9/64) + (3)^2(1/64)$$

$$= 27/64 + 36/64 + 9/64 = 72/64$$

ومنه فإنّ التباين للمتغير X يكون: $\sigma_x^2 = Var(X) = E(X^2) - \mu^2$

$$= 72/64 - (3/4)^2 = 72/64 - 9/16 = 36/64 = 9/16$$

حيث نجد أنّ الانحراف المعياري هو: $\sigma_x = \sqrt{\frac{9}{16}} = \frac{3}{4} = 0.75$

ملاحظة ④: وجدنا أن التباين لمتغير عشوائي X يُعطى بالعلاقة:

$$\sigma_x^2 = Var(X) = E(X - \mu)^2$$

حيث يمثل μ المتوسط الحسابي لهذا المتغير. ولتوضيح العلاقة السابقة نفرض X متغيراً عشوائياً منقطعاً متوسطه الحسابي μ. عندئذٍ يعطى التباين لقيم X (أو لـ X) بالعلاقة:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

وعندما يكون التوزيع لقيم X فئوياً نجد أنّ التباين يساوي:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \mu)^2 f(x_i)$$

حيث تمثل $f_i = f(x_i)$ تكرار المركز x_i $(i = 1,2,\ldots,k)$. ومنه يمكن كتابة التباين بالشكل:

$$\sigma^2=\sum_{i=1}^{k}x_i^2\frac{f(x_i)}{n}-2\mu\sum_{i=1}^{k}x_i\frac{f(x_i)}{n}+\mu^2\sum_{i=1}^{k}\frac{f(x_i)}{n}$$

حيث تمثل $\frac{f(x_i)}{n}$ التكرار النسبي للقيمة x_i $(i=1,2,\ldots,k)$ وهو نفسه قيمة دالة الكثافة الاحتمالية للمتغير العشوائي X عند النقطة x_i أي أنّ:

$$\frac{f(x_i)}{n}=P(X=x_i)=P(x_i)$$

ومنه فإنّ:

$$\begin{aligned}\sigma^2&=\sum_{i=1}^{k}x_i^2P(x_i)-2\mu\sum_{i=1}^{k}x_iP(x_i)+\mu^2\sum_{i=1}^{k}P(x_i)\\&=E(X^2)-2\mu E(X)+\mu^2(1)\\&=E(X^2)-2\mu^2+\mu^2\\&=E(X^2)-\mu^2=Var(X)\end{aligned}$$

3-5-6 خواص التباين لمتغير عشوائي:

بالاعتماد على خواص التوقع الرياضي وعلى تعريف تباين متغير عشوائي نستطيع أن نستنتج خواص التباين التالية:

إذا كانت $u(x)$ دالة في المتغير العشوائي X، وكان k,ℓ عددين ثابتين حيث $k\neq 0$ فإنّ:

1) $$Var\,k=0 \qquad \textbf{(3.38)}$$

2) $$Var\ \ (ku(X)+\ell)=k^2Var(u(X)) \qquad \textbf{(3.39)}$$

3) إذا كانت $u_1(x_1),u_2(x_2),\ldots,u_n(x_n)$ دوال عددية في المتغيرات العشوائية المستقلة $X_1,X_2,\ldots,X_n$ على الترتيب فإنّ:

$$Var(u_1+u_2+\ldots+u_n)=Var(u_1)+Var(u_2)+\ldots+Var(u_n) \qquad \textbf{(3.40)}$$

من الخواص السابقة وفي الحالة الخاصة التي يكون فيها $u(X)=X$ و $u_1(X_1)=X_1,\ldots,u_n(X_n)=X_n$ وبفرض أنّ k عدد ثابت لا يساوي الصفر، نستنتج أنّ:

$$Var(kX+\ell)=k^2Var(X) \qquad \textbf{(3.41)}$$

وإذا كانت $X_1,X_2,\ldots,X_n$ متغيرات عشوائية مستقلة فإنّ:

$$Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n) \qquad \textbf{(3.42)}$$

مثال (23): ليكن X و Y متغيرين عشوائيين مستقلين بحيث أنّ:

$E(X) = 1$ ، $E(Y) = 2$ ، $Var(X) = 2$ ، $Var(Y) = 4$ عندها نجد أنّ:

$$E(5X - 2Y - 3) = 5E(X) - 2E(Y) - 3 = 5 - 4 - 3 = -2$$

$$Var(5X - 2Y - 3) = Var(5X) + Var(-2Y) + Var(-3)$$
$$= 25Var(X) + 4Var(Y) + 0 = 25(2) + 4(4) = 66$$

3-5-7 التباين المزدوج لمتغيرين عشوائيين (The covariance):

يُعرّف التباين المزدوج لمتغيرين عشوائيين X و Y (أو التغاير بين المتغيرين العشوائيين X و Y) بأنه القيمة المتوقعة لحاصل ضرب فرق كل منهما عن متوسطه الحسابي أي أنّ:

$$Cov(X, Y) = \sigma_{xy} = E[(X - \mu_1)(Y - \mu_2)] \qquad \textbf{(3.43)}$$

حيث $\mu_1 = E(X)$ و $\mu_2 = E(Y)$ هما المتوسطان الحسابيان للمتغيرين X و Y على الترتيب.

من العلاقة (3.43) وبالاستفادة من خواص التوقع الرياضي نجد أنّ:

$$Cov(X, Y) = E[(X - \mu_1)(Y - \mu_2)] = E(XY - \mu_1 y - \mu_2 X + \mu_1\mu_2)$$
$$= E(XY) - \mu_1 E(Y) - \mu_2 E(X) + \mu_1\mu_2$$
$$= E(XY) - 2\mu_1\mu_2 + \mu_1\mu_2$$

ومنه:

$$Cov(X, Y) = E(XY) - \mu_1\mu_2 \qquad \textbf{(3.44)}$$

من العلاقة (3.44) الأخيرة نستنتج العلاقة الهامة التالية والتي تعطي القيمة المتوقعة لحاصل ضرب متغيرين عشوائيين:

$$E(XY) = Cov(X, Y) + \mu_1\mu_2 \qquad \textbf{(3.45)}$$

والتي تكتب أيضاً بالشكل التالي:

$$E(XY) = Cov(X, Y) + E(X)E(Y) \qquad \textbf{(3.46)}$$

أي أنّ القيمة المتوقعة لحاصل ضرب متغيرين عشوائيين تساوي حاصل ضرب قيمتهما المتوقعتين مضافاً إليها تباينهما المزدوج.

نتيجة (5):

يمكن بسهولة إثبات أنه إذا كان X و Y متغيرين عشوائيين توقعهما μ_1 و μ_2 على الترتيب وكانت a_1, a_2, b_1, b_2 أعداداً حقيقية حيث $a_1 \neq 0$ و $a_2 \neq 0$ فإنّ:

$$Cov(a_1X + b_1, a_2Y + b_2) = a_1a_2Cov(X, Y) \qquad \textbf{(3.47)}$$

3-5-8 معامل الارتباط بين متغيرين عشوائيين (correlation coefficient):

يُعرف معامل الارتباط بين متغيرين عشوائيين بالعلاقة التالية:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_x.\sigma_y} \qquad \textbf{(3.48)}$$

حيث تمثل σ_x، σ_y الانحرافان المعياريان للمتغيرين X و Y على الترتيب.

ملاحظة ⑤: إنّ معامل الارتباط بين متغيرين عشوائيين هو عدد حقيقي محصور بين العدد (-1) والعدد (1) أي أنّ: $-1 \leq \rho(X, Y) \leq 1$

نتيجة (6):

يمكن بسهولة إثبات أنّ:

$$\rho(a_1X + b_1, a_2Y + b_2) = \rho(X, Y) \qquad \textbf{(3.49)}$$

حيث a_1, a_2, b_1, b_2 أعداد حقيقية و $a_1 \neq 0$, $a_2 \neq 0$.

مثال (24): أوجد التباين المزدوج للمتغيرين العشوائيين المنقطعين X و Y المعرفين في المثال (7)، ثم أوجد معامل الارتباط بين هذين المتغيرين.

الحل: وجدنا في المثال (7) أنّ قانون التوزيع المشترك للمتغيرين X و Y يعطى بالجدول التالي:

X \ Y	0	1	2	المجموع
0	$6/36$	$8/36$	$1/36$	$15/36$
1	$12/36$	$6/36$	0	$18/36$
2	$3/36$	0	0	$3/36$
المجموع	$21/36$	$14/36$	$1/36$	1

وقانون توزيع المتغير العشوائي X يعطى بالجدول:

x	0	1	2
f(x)	$15/36$	$18/36$	$3/36$

وقانون توزيع المتغير العشوائي Y يعطى بالجدول:

y	0	1	2
g(y)	$21/36$	$14/36$	$1/36$

لإيجاد التباين المزدوج للمتغيرين X و Y نحتاج لحساب التوقع الرياضي لكل من X و Y و XY حيث نجد أنّ:

$$E(X)=\sum_x xf(x)=\frac{2}{3}\ ,\ E(Y)=\sum_y yg(y)=\frac{4}{9}$$

$$E(XY)=\sum_x\sum_y xy\quad f(x,y)=\frac{1}{6}$$

ومنه فإنّ التباين المزدوج للمتغيرين X و Y يكون:

$$\begin{aligned}Cov(X,Y)&=E(XY)-E(X)E(Y)\\&=\frac{1}{6}-\frac{2}{3}.\frac{4}{9}=-\frac{7}{54}=-0.129\end{aligned}$$

ولإيجاد معامل الارتباط بين المتغيرين العشوائيين X و Y نحتاج لحساب σ_x، σ_y حيث نجد أنّ:

$$E(X^2)=\sum_x x^2f(x)=\frac{10}{12}\ ,\ E(Y^2)=\sum_y y^2g(y)=\frac{1}{2}$$

ومنه:

$$\begin{aligned}\sigma_x^2&=E(X^2)-[E(X)]^2\\&=\frac{10}{12}-\frac{4}{9}=\frac{14}{36}=0.388\\\sigma_y^2&=E(Y^2)-[E(Y)]^2\\&=\frac{1}{2}-\frac{16}{81}=\frac{98}{324}=0.302\end{aligned}$$

وبالتالي فإنّ معامل الارتباط بين المتغيرين X و Y يكون:

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sigma_x\quad\sigma_y}=\frac{-0.129}{\sqrt{0.388}\sqrt{0.302}}=-0.378$$

مثال (25): أوجد التباين المزدوج للمتغيرين العشوائيين المنقطعين X و Y المعرفين في المثال (10)، ثم أوجد معامل الارتباط بينهما.

الحل: وجدنا في المثال (10) أنّ قانون التوزيع المشترك للمتغيرين X و Y هو:

X \ Y	1	2	3
1	$1/9$	$1/9$	$1/9$
2	$1/9$	$1/9$	$1/9$
3	$1/9$	$1/9$	$1/9$

كما وجدنا أنّ قانون توزيع المتغير العشوائي X هو:

x	1	2	3
f(x)	$1/3$	$1/3$	$1/3$

وقانون توزيع المتغير العشوائي Y هو:

y	1	2	3
g(y)	$1/3$	$1/3$	$1/3$

ومنه نجد أنّ:

$$E(X) = \sum_x x f(x) = 2$$

$$E(Y) = \sum_y y g(y) = 2$$

$$E(XY) = \sum_x \sum_y xy \quad f(x,y) = 4$$

وهكذا فإنّ التباين المزدوج للمتغيرين X و Y يكون:

$$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= 4 - (2)(2) = 0 \end{aligned}$$

وبالتالي فإنّ معامل الارتباط بين المتغيرين X و Y يكون:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_x \quad \sigma_y} = 0$$

لأنّ:

$$Cov(X,Y) = 0$$

لاحظ بأن التباين المزدوج للمتغيرين X و Y ومعامل الارتباط بينهما معدومان في المثال السابق وذلك لأنّ X و Y متغيران عشوائيان مستقلان في ذلك المثال.

نتيجة (7):

إذا كان X و Y متغيرين عشوائيين مستقلين فإنّ: $Cov(X, Y) = 0$ وكذلك $\rho(X, Y) = 0$

نتيجة (8):

يمكن تعميم مفهوم التباين المزدوج لمتغيرين عشوائيين إلى التباين المتعدد لثلاثة متغيرات عشوائية أو أكثر حيث نجد أنّ التباين المتعدد لثلاثة متغيرات عشوائية Z, Y, X هو:

$$Cov(X, Y, Z) = E[(X-\mu_1)(Y-\mu_2)(Z-\mu_3)] \qquad \textbf{(3.50)}$$

حيث يمثل μ_3, μ_2, μ_1 المتوسطات الحسابية للمتغيرات Z, Y, X على الترتيب. ويمكننا الإثبات بسهولة أنّ:

$$\begin{aligned} Cov(X, Y, Z) = {} & E(XYZ) - \mu_1\mu_2\mu_3 - \mu_1 E[(Y-\mu_2)(Z-\mu_3)] \\ & -\mu_2 E[(X-\mu_1)(Z-\mu_3)] - \mu_3 E[(X-\mu_1)(Y-\mu_2)] \end{aligned} \qquad \textbf{(3.51)}$$

ومن العلاقة الأخيرة نجد أنّ:

$$\begin{aligned} E(XYZ) = {} & Cov(X, Y, Z) + \mu_1\mu_2\mu_3 + \mu_1 E[(Y-\mu_2)(Z-\mu_3)] \\ & +\mu_2 E[(X-\mu_1)(Z-\mu_3)] + \mu_3 E[(X-\mu_1)(Y-\mu_2)] \end{aligned} \qquad \textbf{(3.52)}$$

كذلك يمكن تعميم مفهوم معامل الارتباط بين متغيرين عشوائيين ليشمل عدداً من المتغيرات العشوائية. حيث نجد أن معامل الارتباط بين ثلاثة متغيرات عشوائية Z, Y, X هو:

$$\rho(X, Y, Z) = \frac{Cov(X, Y, Z)}{\sigma_x \ \sigma_y \ \sigma_Z} \qquad \textbf{(3.53)}$$

3-6 العزوم والدوال المولدة لها
:(Moments and Moment Generating functions)

يتميز كل من المتوسط الحسابي $\mu = E(X)$ والتباين $\sigma^2 = E(X-\mu)^2$ لمتغير عشوائي مثل X بأنهما من المقاييس الوصفية العددية التي تحدد مركزية التوزيعات التكرارية وتراكمهـا أو تجمعها حول نقطة أو مفهوم معين. كما تصف أيضاً انتشار التوزيع الاحتمالي للمتغير P(X) ضمن نطاق معين، ولكن هاتين الدالتين، المتوسط والتباين غير كافيتين لتحديد خاصية التوزيع وذلك لأن هنالك الكثير من التوزيعات التي قد تشترك بنفس المتوسط أو بنفس التبـاين علـى الرغم من اختلافها الجوهري.

ومن المقاييس الوصفية الكمية الأخرى التي يمكن استخدامها لتحديـد خصوصـية التوزيـع التكراري وتحديد احتمالاته P(X) بصورة موحدة هو ما يمكن تسميته بالعزوم حـول قيمـة معينة.

3-6-1 تعريف العزم:

يُعرف العزم الكائي (من المرتبة k) لمتغير عشوائي مثل X حول المتوسط الحسابي μ لهـذا المتغير ويُرمز له بالرمز M_k، (أو حول أي قيمة معينة أخرى لهـذا المتغيـر مثـل $x = 0$ ويُرمز له عندئذٍ بالرمز M'_k)، بأنه القيمة المتوقعة للقوة k لانحراف المتغيـر عـن وسـطه الحسابي (أو لانحراف المتغير عن تلك القيمة المعينة له). ومنه فإنّ العـزم الكـائي لمتغيـر عشوائي منقطع X حول متوسطه الحسابي μ هو:

$$M_k = E[(X-\mu)^k] = \sum_{i=1}^{\infty}(x_i-\mu)^k f(x_i) \qquad (3.54)$$

والعزم لهذا المتغير X حول نقطة الأصل $x = 0$ هو:

$$M'_k = E(X^k) = \sum_{i=1}^{\infty} x_i^k f(x_i) \qquad (3.55)$$

وإذا كان المتغير العشوائي X مستمراً فإنّ:

$$M_k = E[(X-\mu)^k] = \int_{-\infty}^{\infty}(x-\mu)^k f(x)dx \qquad (3.56)$$

$$M'_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx \qquad (3.57)$$

من التعريف السابق للعزم نلاحظ بأنه هنالك ثلاث حالات مميزة وهي:

أ – عندما يكون المتوسط الحسابي $\mu = 0$ (أو القيمة المعينة للمتغير العشوائي هي نقطة الأصل)، وكذلك عندما يكون $k = 1$ فإنّ العزم الأول للمتغير العشوائي X هو في الواقع القيمة المتوقعة أو المتوسط الحسابي لهذا المتغير أي أنّ:

$$M'_1 = E(X) = \mu \qquad (3.58)$$

ب– عندما $k = 2$ و $\mu \neq 0$ فإنّ العزم الثاني للمتغير العشوائي X حول μ هو في الواقع التباين لهذا المتغير أي أنّ:

$$M_2 = E[(X-\mu)^2] = \sigma^2 \qquad (3.59)$$

حيث نلاحظ في هذه الحالة أنّ: $M_2 = \sigma^2 = M'_2 - (M'_1)^2$

لأنّ: $M'_2 = E(X^2)\ ,\ M'_1 = E(X)$

جـ– إذا كانت كافة العزوم لمتغيرين عشوائيين مثل Y, X حول نقطة الأصل متساوية أي كانت: $(M'_1)_x = (M'_1)_y, \ldots, (M'_k)_x = (M'_k)_y$

فإنّ ذلك يعني بأنّ التوزيع التكراري $P(X)$ للمتغير X يساوي تقريباً أو يشابه أو يؤول إلى التوزيع التكراري $P(Y)$ للمتغير Y. وهذا مالا يمكن استنتاجه من تساوي المتوسط الحسابي والتباين لكل من المتغيرين فقط.

3-6-2 تعريف ا لدالة المولدة للعزوم (Moment Generating function):

تُعرّف الدالة المولدة لعزوم متغير عشوائي مثل X، بأنها القيمة المتوقعة للدالة الأسية لمضاعفات المتغير X ويرمز لها بالرمز $m_x(t)$. وبتعبير رياضي تُعرّف الدالة المولدة للعزوم بالعلاقة التالية:

$$m(t) = E(e^{tX}) \qquad (3.60)$$

ويتحدد منطلق $m(t)$ بحيث يكون التوقع في الطرف الثاني موجوداً.

إذا كان X متغيراً عشوائياً منقطعاً وإذا افترضنا بأنّ العزوم الكائية M'_k حول نقطة الأصل موجودة فإننا نجد بأنّ الدالة المولدة للعزوم للمتغير العشوائي X تأخذ الشكل التالي عند تواجدها:

$$m(t) = E(e^{tX}) = \sum_x e^{tx} f(x)$$

$$= \sum_x \left(1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \ldots\right) f(x)$$

$$=\sum_x f(x)+t\sum_x xf(x)+\frac{t^2}{2!}\sum_x x^2f(x)+\ldots$$

ومنه:

$$m(t)=1+tM'_1+\frac{t^2}{2!}M'_2+\frac{t^3}{3!}M'_3+\frac{t^4}{4!}M'_4+\ldots \qquad \textbf{(3.61)}$$

وهكذا نستنتج من العلاقة الأخيرة بأنّ الدالة المولدة للعزوم $m(t)$ هي عبارة عن دالـــة تضـــم كافة العزوم حول نقطة الأصل ولهذا سميت بالدالة المولدة حيث يمكن استخلاص أو إيجاد أي عزم من العزوم وذلك بتكرار عملية الاشتقاق للدالة المولدة كما سنجد في النظرية التالية:

نظرية ❷ :

إذا كانت الدالة المولدة للعزوم $m(t)$ لمتغير عشوائي X موجودة فإنّ:

$$\left.\frac{d^k m(t)}{dt^k}\right|_{t=0}=M'_k \qquad \textbf{(3.62)}$$

الإثبات:

بما أنّ القيمة الحقيقية للدالة المولدة $m(t)$ موجودة، لهذا بأخذ المشتق بصورة متتالية للعلاقـــة (3.61) ومن ثم وضع $t=0$ في كل مرة نحصل على العزوم المطلوبة. فإذا افترضنا أنّ X متغير عشوائي منقطع وبأخذ المشتق الأول للعلاقة (3.61) نجد أنّ:

$$\frac{d}{dt}m(t)=0+M'_1+tM'_2+\frac{t^2}{2!}M'_3+\frac{t^3}{3!}M'_4+\ldots$$

وبوضع $t=0$ يكون:

$$\left.\frac{d}{dt}m(t)\right|_{t=0}=M'_1$$

وإذا أخذنا المشتق الثاني للعلاقة (3.61) نجد أنّ:

$$\frac{d^2}{dt^2}m(t)=M'_2+tM'_3+\frac{t^2}{2!}M'_4+\ldots$$

وبوضع $t=0$ يكون:

$$\left.\frac{d^2}{dt^2}m(t)\right|_{t=0}=M'_2$$

وهكذا عند أخذ المشتق من المرتبة k للعلاقة (3.61) والتعويض بـ $t=0$ نجد أنّ:

$$\left.\frac{d^k}{dt^k}m(t)\right|_{t=0}=M'_k$$

ملاحظة ⑥ : إذا كان X متغيراً عشوائياً مستمراً فإنّ الدالة المولدة للعزوم لهذا المتغير تُعرف بالتكامل التالي:

$$m(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \qquad \textbf{(3.63)}$$

حيث نجد أنّ:

$$m(t) = \int_{-\infty}^{\infty} (1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \ldots) f(x)dx$$
$$= \int_{-\infty}^{\infty} f(x)dx + \int_{-\infty}^{\infty} txf(x)dx + \int_{-\infty}^{\infty} \frac{(tx)^2}{2!} f(x)dx + \ldots$$
$$= 1 + tM_1' + \frac{1}{2!}t^2M_2' + \frac{t^1}{3!}t^3M_3' + \ldots$$

وهي نفس العلاقة (3.61) التي وجدناها في حالة المتغير العشوائي المنقطع.

ملاحظة ⑦ : إذا كان X متغيراً عشوائياً منقطعاً أو مستمراً فإننا سنرمز للتوضيح عند الضرورة للدالة المولدة لعزوم هذا المتغير بالرمز $m_x(t)$. وبصورة عامة إذا كانت $u(x)$ دالة في المتغير العشوائي X فإنه يمكننا تعريف الدالة المولدة لعزوم $u(X)$ كما يلي:

في الحالة التي يكون فيها المتغير العشوائي X منقطعاً:

$$m_{u(x)}(t) = E(e^{tu(X)}) = \sum_x e^{tu(x)} f(x) \qquad \textbf{(3.64)}$$

وفي الحالة التي يكون فيها المتغير العشوائي X مستمراً نجد أنّ:

$$m_{u(x)}(t) = E(e^{tu(X)}) = \int_{-\infty}^{\infty} e^{tu(x)} f(x)dx \qquad \textbf{(3.65)}$$

3-6-3 خواص الدالة المولدة للعزوم:

إنّ للدالة المولدة للعزوم خصائص نذكر منها الخواص التالية:

إذا كانت $u(x)$ دالة في المتغير العشوائي X و k ثابت ما فإنّ:

1) $$m_{ku(x)}(t) = E(e^{tku(X)}) = m_{u(x)}(kt) \qquad \textbf{(3.66)}$$

2) $$m_{u(x)+k}(t) = E(e^{t(u(X)+k)}) = e^{tk} m_{u(x)}(t) \qquad \textbf{(3.67)}$$

3) إذا كان المتغيران العشوائيان Y,X مستقلين، وكانت $u(x)$ و$v(y)$ دالتين في Y,X على الترتيب فإنّ:

$$m_{u(x)+v(y)}(t) = E(e^{t(u(X)+v(Y))}) = E(e^{tu(X)}.e^{tv(Y)})$$
$$= E(e^{tu(X)}).E(e^{tv(Y)})$$

ومنه:

$$m_{u(x)+v(y)}(t) = m_{u(x)}(t).m_{v(y)}(t) \qquad \textbf{(3.68)}$$

وفي الحالة الخاصة التي يكون فيها $u(X) = X$ و $v(Y) = Y$ تصبح الخواص السابقة بالشكل:

1) $$m_{kx}(t) = E(e^{tkX}) = m_x(kt) \qquad \textbf{(3.69)}$$

2) $$m_{x+k}(t) = E(e^{t(X+k)}) = e^{tk}m_x(t) \qquad \textbf{(3.70)}$$

حيث k ثابتة ما.

3) إذا كان المتغيران العشوائيان Y, X مستقلين فإنّ:

$$m_{x+y}(t) = E(e^{t(X+Y)}) = E(e^{tX}.e^{tY}) = E(e^{tX}).E(e^{tY})$$

ومنه:

$$m_{x+y}(t) = m_x(t).m_y(t) \qquad \textbf{(3.71)}$$

ويمكن تعميم الخاصة الأخيرة كما يلي: إذا كانت $X_1, X_2, \ldots, X_n$ متغيرات عشوائية مستقلة فإنّ:

$$M_{x_1+x_2+\ldots+x_n}(t) = M_{x_1}(t).M_{x_2}(t)\ldots M_{x_n}(t) \qquad \textbf{(3.72)}$$

وإذا كان لهذه المتغيرات نفس القانون إضافة إلى كونها مستقلة أي إذا كان: $X_1 = X_2 = X_3 = \ldots = X_n$ فإنّ:

$$M_{x_1+x_2+\ldots+x_n}(t) = M_{nx_1}(t) = [M_{x_1}(t)]^n \qquad \textbf{(3.73)}$$

مثال (26): ليكن X المتغير العشوائي الثنائي (أو الحداني) وهو متغير منقطع دالة كثافته:

$$f(x) = P(X = x) = \binom{n}{x} P^x q^{n-x} \qquad x = 0,1,2,\ldots,n$$

ووسيطاه n, p حيث تمثل n عدد المحاولات في تجربة ثنائية (التجربة التي تكون نتيجتها النجاح أو الفشل) و p احتمال النجاح أما q فهي تمثل احتمال الفشل وتساوي $q = 1 - p$

عندئذٍ نجد أنّ الدالة المولدة لعزوم هذا المتغير هي:

$$m_x(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} f(x)$$

$$= \sum_{x=0}^{n} e^{tx} \binom{n}{x} P^x q^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x}$$

$$= (pe^t + q)^n$$

وذلك أياً كان t من R.

مثال (27): ليكن X المتغير العشوائي البواسوني وهو متغير منقطع دالة كثافته:

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots$$

ووسيطه λ يمثل معدل عدد مرات النجاح في تجربة ثنائية.

عندئذٍ نجد أنّ الدالة المولدة لعزوم هذا المتغير هي:

$$m_x(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} f(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} = e^{-\lambda} e^{\lambda e^t}$$

وذلك أياً كان t من R.

مثال (28): ليكن X المتغير العشوائي النظامي (أو الطبيعي) وهو متغير عشوائي مستمر وسيطاه μ و σ^2 ودالة كثافته هي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad , \qquad -\infty < x < \infty$$

عندئذٍ نجد أنّ الدالة المولدة لعزوم هذا المتغير هي:

$$m_x(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

$$= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

وبفرض $y = x - \mu$ يكون $dy = dx$ وتبقى حدود التكامل السابق من $-\infty$ إلى $+\infty$ حيث نجد أنّ:

$$m_x(t) = \int_{-\infty}^{\infty} e^{t(y+\mu)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} dy$$

$$=e^{\mu t}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{ty-(\frac{y^2}{2\sigma^2})}dy$$

فإذا لاحظنا أنّ:
$$ty-\frac{y^2}{2\sigma^2}=-\frac{(y-\sigma^2t)^2}{2\sigma^2}+\frac{\sigma^2t^2}{2}$$

فإننا نستطيع أن نكتب:
$$m_x(t)=e^{\mu t}e^{\frac{\sigma^2t^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\sigma^2t)^2}{2\sigma^2}}dy=e^{\mu t+\frac{\sigma^2t^2}{2}}$$

وذلك أياً كان t من R.

مثال (29): إذا كانت الدالة المولدة لعزوم المتغير العشوائي المستمر X معرفة كالتالي:

$$m(t)=\begin{cases}\frac{1}{(1-t)^2} & , \quad t<1\\ 0 & , \quad \text{عدا ذلك}\end{cases}$$

فأحسب كلاً من المتوسط الحسابي والتباين للمتغير X.

الحل: بالاستفادة من النظرية ❷ نجد أنّ:

$$\left.\frac{dm(t)}{dt}\right|_{t=0}=\left.\frac{2}{(1-t)^3}\right|_{t=0}=2=M_1'$$

$$\left.\frac{d^2m(t)}{dt^2}\right|_{t=0}=\left.\frac{6}{(1-t)^4}\right|_{t=0}=6=M_2'$$

وبما أنّ: $\mu=M_1'$ فإنّ المتوسط الحسابي للمتغير X هو: $\mu=2$

وكذلك بما أنّ: $\sigma^2=M_2'-(M_1')^2$ فإنّ التباين للمتغير X هو: $\sigma^2=6-4=2$

مثال (30): إذا كانت دالة الكثافة الاحتمالية لمتغير عشوائي منقطع X معرفة كالتالي:

$$f(x)=\begin{cases}\frac{6}{\pi^2x^2} & , \quad x=1,2,3,\ldots\\ 0 & \quad \text{عدا ذلك}\end{cases}$$

بيّن بأنّ الدالة المولدة لعزوم هذا المتغير غير موجودة أو غير معرفة.

الحل: بما أنّ المتغير X غير مستمر لذا فالدالة المولدة للعزوم (إن وجدت) تحسب من العلاقة التالية:

$$m_x(t) = E(e^{tX}) = \sum_x e^{tx} f(x) = \sum_{x=1}^{\infty} \frac{6e^{tx}}{\pi^2 x^2}$$

ولتوضيح ما إذا كانت هذه المتسلسلة اللانهائية متقاربة أو متباعدة نستخدم اختبار دالمبير كما يلي:

$$\frac{u_{x+1}}{u_x} = \frac{6e^{t(x+1)} / \pi^2 (x+1)^2}{6e^{tx} / \pi^2 x^2} = \frac{x^2 e^t}{(x+1)^2}$$

ومنه:

$$\lim_{x\to\infty} \frac{u_{x+1}}{u_x} = e^t \lim_{x\to\infty} \left(\frac{x}{x+1}\right)^2 = e^t$$

وبما أنّ هذه النهاية أكبر من الواحد عندما $t > 0$ لذا فالمتسلسلة غير متقاربة أي متباعدة وهذا يعني عدم وجود قيمة لهذه المتسلسلة أي أنّ الدالة المولــدة للعــزوم $m_x(t)$ لهــذا المتغيــر العشوائي غير موجودة أو غير معرفة. أو نقول بأنّ المتغير X لا يمتلك دالة مولدة لعزومه.

3-7 تمارين عامة:

1- في كل حالة من الحالات التالية، وضّح ما إذا كانت القاعدة X المعطــاة تعتبــر متغيــراً عشوائياً أم لا مع ذكر الأسباب.

أ – عند رمي حجر نرد مرة واحدة، والقاعدة X معرفة كالآتي:

$$X=\begin{cases} 1 & \text{إذا كانت نتيجة التجربة عدداً زوجياً} \\ -1 & \text{إذا كانت نتيجة التجربة عدداً فردياً} \end{cases}$$

ب – عند رمي قطعة نقود مرة واحدة، والقاعدة X معرفة كالآتي:

$$X=\begin{cases} 10 & \text{إذا كانت النتيجة صورة} \\ 5 & \text{إذا كانت النتيجة كتابة} \\ 9 & \text{إذا كانت النتيجة صورة أو كتابة} \end{cases}$$

جـ – عند رمي قطعة نقود مرة واحدة والقاعدة X معرفة كالآتي:

$$X=\begin{cases} 1 & \text{إذا كانت النتيجة صورة} \\ -1 & \text{إذا كانت النتيجة كتابة} \end{cases}$$

د – عند رمي حجر نرد مرة واحدة، والقاعدة X معرفة كالآتي:

$$X=\begin{cases} 1 & \text{إذا كانت نتيجة التجربة عدداً زوجياً} \\ 0 & \text{إذا كانت نتبجة التجربة العدد 5} \end{cases}$$

2- في كل حالة من الحالات التالية وضّح ما إذا كان المتغير العشوائي منقطعاً أم متصلاً:

أ – المتغير العشوائي يدّل على عدد السيارات التي تم صيانتها في ورشة ما في يوم ما.

ب – المتغير العشوائي يدلُ على كمية هطول الأمطار في دمشق في يوم ما من أيــام فصــل الشتاء.

جـ – المتغير العشوائي يدلُ على عدد الإجابات الصحيحة التي يحصل عليها طالب مــا فــي امتحان مؤلف من 20 سؤالاً.

د– الزمن الذي يأخذه ميكانيكي سيارات لتركيب جهاز العادم لسيارة ما.

3- إذا كان X متغيراً عشوائياً يدلُ على مجموع العددين الظاهرين عند رمي حجر نرد مرتين متتاليتين:

أ – أكتب جميع القيّم التي يمكن أن يأخذها المتغير X

ب– أوجد الحوادث المناظرة لكل قيمة من قيم X في أ.

4- إذا رميت قطعة نقود ثلاث مرّات متتالية، وإذا كان المتغير X يرمز إلى الفرق بين عدد الصور وعدد الكتابات في الثلاث رميات المذكورة.

أ – أكتب القيم الممكنة للمتغير العشوائي X.

ب – أوجد الحوادث المناظرة لكل قيمة من قيم X المختلفة.

جـ – أوجد الاحتمال لكل حدث من هذه الحوادث.

5- في كل حالة من الحالات التالية، وضّح ما إذا كانت الدالة المعطاة يمكن اعتبارها دالة كثافة احتمالية لمتغير عشوائي ما أم لا مع ذكر الأسباب:

(a)
$$f(x) = \begin{cases} 0.5 & , \quad \text{عندما } x=3 \\ 0.2 & , \quad \text{عندما } x=4 \\ 0.3 & , \quad \text{عندما } x=5 \end{cases}$$

(b)
$$f(x) = \begin{cases} -0.3 & , \quad \text{عندما } x=1 \\ 0.7 & , \quad \text{عندما } x=2 \\ 0.6 & , \quad \text{عندما } x=3 \end{cases}$$

(c)
$$f(x) = \begin{cases} 1.2 & , \quad \text{عندما } x=4 \\ 0.6 & , \quad \text{عندما } x=6 \end{cases}$$

(d)
$$f(x) = \begin{cases} 1 & , \quad 0 \le x \le 2 \\ 0 & , \quad \text{عدا ذلك} \end{cases}$$

$$f(x) = \begin{cases} x & , \quad -\frac{1}{2} \le x \le \frac{1}{2} \\ 0 & , \quad \text{عدا ذلك} \end{cases} \qquad (e)$$

$$f(x) = \begin{cases} x & , \quad 0 \le x \le 1 \\ 0 & , \quad \text{عدا ذلك} \end{cases} \qquad (f)$$

6- في تجربة رمي حجر النرد. ليكن X المتغير العشوائي الدال على العدد الأكبر من الوجهين الظاهرين، وعلى أحدهما في حال تساويهما. أكتب جدول التوزيع الاحتمالي للمتغير العشوائي X وأحسب التوقع الرياضي له.

7- كيس يحوي 5 كرات بيضاء و 3 خضراء. سحبنا 4 كرات عشوائياً فإذا كان المتغير العشوائي Y يدل على عدد الكرات البيضاء المسحوبة، أكتب جدول التوزيع الاحتمالي للمتغير العشوائي Y وأحسب توقعه الرياضي وتباينه.

8- مجموعة مكونة من 8 طلاب 3 منهم من قسم هندسة الغزل والنسيج. أخترنا ثلاثة طلاب من المجموعة بصورة عشوائية. بفرض أنّ المتغير العشوائي X يدل على عدد طلاب قسم الغزل والنسيج في المجموعة المختارة فالمطلوب:

1- أكتب جدول التوزيع الاحتمالي للمتغير العشوائي X.

2- أحسب التوقع الرياضي والتباين والانحراف المعياري للمتغير العشوائي X.

9- شركة فيها 11 مهندساً 3 منهم اختصاص طاقة كهربائية. أرادت الشركة تشكيل لجنة مؤلفة من 5 مهندسين بصورة عشوائية. فإذا كان المتغير العشوائي X يدل على عدد مهندسي الطاقة الكهربائية في المجموعة المختارة فالمطلوب:

1- أكتب جدول التوزيع الاحتمالي للمتغير العشوائي X.

2- أحسب التوقع الرياضي والتباين والانحراف المعياري للمتغير العشوائي X.

10- ليكن X متغيراً عشوائياً منفصلاً له جدول التوزيع الاحتمالي التالي:

x	0	1	2	3
$p(X=x)$	0.1	0.35	0.4	0.15

والمطلوب:

1- أحسب التوقع الرياضي للمتغير العشوائي X. ثم أحسب التباين والانحراف المعياري لهذا المتغير.

2- أحسب التوقع الرياضي للدالة: $g(x)=1+2\sin x$

11- أراد شخص إرسال ثلاث رسائل إلى أصدقائه. كتب الرسائل والعناوين على الظروف ونسي أن يضع كل رسالة في الظرف المخصص لها. جاء أبنه ووضع عشوائياً كل رسالة في ظرف ثم وضع الرسائل في البريد. أكتب فضاء التجربة. ثم أحسب احتمال وصول كل رسالة إلى صاحبها الصحيح، واحتمال عدم وصول أي رسالة لصاحبها وأخيراً احتمال وصول رسالة واحدة فقط لصاحبها الصحيح.

12- يريد أحد رجال الأعمال حساب القيمة المتوقعة (المتوسط الحسابي) لعوائد استثماراته، فإذا كان X المتغير العشوائي الدال على كمية هذه العوائد، وكانت الدالة الاحتمالية للمتغير X هي:

$$f(x)=P(X=x)=\begin{cases} 0.9 \quad , \quad x=10{,}000\$ & \text{عندما} \\ 0.1 \quad , \quad x=-1000\$ & \text{عندما} \end{cases}$$

أوجد القيمة المتوقعة لهذه العوائد.

13- ليكن X متغيراً عشوائياً منقطعاً دالة كثافته الاحتمالية معرفة بالشكل:

$$f(x)=P(X=x)=\frac{1}{6} \quad , \quad x=1,2,3,4,5,6$$

أوجد القيمة المتوقعة والإنحراف المعياري للمتغير X.

14- إذا رمينا قطعة نقود متوازنة مرة واحدة، وعرفنا المتغير العشوائي X كما يلي:

$$X=\begin{cases} 1 & ,\quad \text{إذا كانت النتيجة صورة} \\ 3 & ,\quad \text{إذا كانت النتيجة كتابة} \end{cases}$$

أ – أوجد المتوسط الحسابي والتباين للمتغير العشوائي X

ب- إذا وضعنا $Y = X + 5$ ، فما هو المتوسط الحسابي والانحراف المعياري للمتغير الجديد Y.

15- ليكن X متغيراً عشوائياً مستمراً دالة كثافته معرفة بالشكل:

$$f(x) = 6(1-x)x \quad , \quad 0 < x < 1$$

أوجد التوقع الرياضي والتباين والانحراف المعياري لهذا المتغير.

16- متغير عشوائي مستمر X دالة كثافته هي:

$$f(x)=\begin{cases} \frac{1}{2}e^{-\frac{1}{2}x} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

والمطلوب:

1- تأكد أنّ $f(x)$ هي دالة كثافة احتمالية.

2- أوجد دالة التوزيع المجمعة للمتغير العشوائي X.

3- أحسب الاحتمال $P(1 \leq X \leq 2)$

4- أوجد التوقع الرياضي للمتغير العشوائي X وكذلك التباين والانحراف المعياري.

17- متغير عشوائي مستمر X دالة كثافته هي:

$$f(x)=\begin{cases} \sin 2x & , \quad 0 \leq x \leq \frac{\pi}{2} \\ 0 & , \quad \text{عدا ذلك} \end{cases}$$

والمطلوب:

1- تأكد أنّ $f(x)$ هي دالة كثافة احتمالية.

2- أحسب الاحتمال $P(\frac{\pi}{6} \leq X \leq \frac{\pi}{2})$

3- أوجد التوقع الرياضي للمتغير العشوائي X وكذلك التباين والانحراف المعياري.

4- أوجد التوقع الرياضي للدالة $g(x)=\dfrac{3}{1+\cos^2 x}$

18- ليكن X متغيراً عشوائياً مستمراً معرفاً بالدالة:

$$f(x)=\frac{a}{x^2+1}\quad,\quad -\infty<x<\infty$$

حيث a ثابت عددي والمطلوب:

1- عيّن قيمة الثابت a لكي تكون الدالة f(x) دالة كثافة احتمالية للمتغير العشوائي X.

2- أوجد دالة التوزيع المجمعة للمتغير X.

3- أحسب الاحتمال $p(-1\le X\le 1)$

19- Y,X متغيران عشوائيان منقطعان دالة كثافتهما المشتركة معرفة بالجدول التالي:

X \ Y	-1	0	1
0	0.1	0.2	0.1
1	0.2	0.3	0.1

أوجد معامل الارتباط بين Y,X .

20- يحتوي جهاز لا يعمل على أربع وصلات أساسية واحدة منها بيضاء وأثنتان حمراوان وواحدة سوداء. تم فحص ثلاث وصلات بشكل عشوائي. ليكن X المتغير العشوائي الدال على عدد الوصلات البيضاء التي تم فحصها وY المتغير العشوائي الدال على عدد الوصلات الحمراء التي تم فحصها أيضاً. والمطلوب:

1- عيّن جدول التوزيع المشترك للمتغيرين العشوائيين X وY .

2- أوجد معامل الارتباط بين X وY

3- بيّن فيما إذا كان X وY مستقلين أم لا؟

21- ليكن X و Y متغيران عشوائيان مستمران دالة كثافتهما المشتركة معرفة بالعلاقة:

$$f(x,y)=\frac{1}{2\pi}e^{-\frac{x^2-2xy+2y^2}{2}}\quad,\quad -\infty<x,y<\infty$$

والمطلوب:

1- اثبت أنّ الدالة $f(x,y)$ هي دالة كثافة احتمالية.

2- أوجد الكثافتين الهامشيتين لكل من X و Y .

3- أوجد التوقع الرياضي لكل من X و Y .

4- أوجد الانحراف المعياري لكل من X و Y .

5- بيّن فيما إذا كان X و Y مستقلين أم لا؟

6- أوجد التباين المزدوج للمتغيرين X و Y ومعامل الارتباط بينهما.

22- X و Y متغيران عشوائيان مستمران دالة كثافتهما المشتركة معرفة بالعلاقة:

$$f(x,y) = y^2.e^{-y(x+1)} \quad , \quad x \geq 0, y \geq 0$$

والمطلوب:

1- اثبت أنّ الدالة $f(x,y)$ هي دالة كثافة احتمالية.

2- أوجد الكثافتين الهامشيتين لكل من X و Y . ثم بيّن فيما إذا كان المتغيران X و Y مستقلان أم لا؟

3- أوجد التباين المزدوج للمتغيرين العشوائيين X و Y .

4- أوجد الانحراف المعياري لكل من X و Y .

5- أوجد معامل الارتباط بين X و Y

23- اخترنا نقطة بصورة عشوائية على القطر المطابق لمحور السينات من دائرة مركزها المبدأ ونصف قطرها الواحد. لتكن x فاصلة هذه النقطة. ننشئ الوتر العمود على محور السينات في هذه النقطة ثم نختار بصورة عشوائية نقطة من هذا الوتر وليكن y ترتيب هذه النقطة. والمطلوب:

1- أوجد دالة الكثافة المشتركة للمتغيرين المستمرين X و Y

2- ما هو احتمال وقوع النقطة التي نختارها بهذا الشكل داخل المربع المرسوم ضمن الدائرة والذي تتوازى أضلاعه مع المحاور الإحداثية.

24- ليكن X متغيراً عشوائياً بواسونياً وسيطه λ. استخدم الدالة المولدة لعزوم X لحساب التوقع الرياضي والتباين لهذا المتغير.

25- متغير عشوائي مستمر X دالة كثافته الاحتمالية:

$$f(x) = \frac{1}{2}e^{-|x|} \qquad \forall x \in R$$

والمطلوب:

1- أحسب $m_x(t)$.

2- أوجد صيغة تعطي $E(X^{2n})$ مستخدماً الدالة المولدة للعزوم.

26- إذا كانت $X_1, X_2, \ldots, X_n$ متغيرات عشوائية مستقلة ولها قانون التوزيع نفسه وكانت الدالة المولدة لعزوم كل من هذه المتغيرات $m(t)$ منتهية مهما يكن t. أثبت أنّ:

$$E(X_1 + \ldots + X_n)^3 = nE(X_1^3) + 3n(n-1)E(X_1^2)E(X_1) + n(n-1)(n-2)(E(X_1))^3$$

الفصل الرابع

التوزيعات الخاصة

4-1 – التوزيع الثنائي

4-2 – توزيع بواسون

4-3 – التوزيع الطبيعي

4-4 – توزيع (ستيودنت) t.

4-2 – توزيع مربع كاي (χ^2).

4-6 – توزيع فيشر.

4-7 – تمارين عامة.

التوزيعات الخاصة
(Special Distributions)

عرفنا في الفصول السابقة كيفية انتشار التكرارات النسبية والعادية للمتغيرات العشوائية المتصلة منها والمنفصلة من خلال ما يسمى بالمدرجات والمضلعات والمنحنيات التكرارية حيث تأخذ هذه المضلعات والمنحنيات أشكالاً خاصة يطلق عليها أسم التوزيعات التكرارية بينما يطلق على الدوال التي تمثل تلك الأشكال أسم دوال التوزيع الخاصة. وتتعدد دوال التوزيع الشهيرة منها على ضوء استخداماتها وأهميتها في التجارب الإحصائية الميدانية، كما تصنف هذه الدوال إلى نوعين هامين على أساس نوعية المتغير العشوائي إن كان متصلاً (مستمراً) أم منفصلاً (متقطعاً) وهما:

أ – دوال التوزيع المتصلة　　　　ب- دوال التوزيع المنفصلة

ونذكر فيما يلي بعضاً من كلا النوعين المذكورين:

4-1 التوزيع الثنائي (Binomial distribution):

يطلق أسم التوزيع الثنائي على حصيلة التجارب ذات المحاولات المتكررة والمستقلة التي تحتوي كل منها على نتيجتين فقط مثل النجاح أو الفشل، الصورة أو الكتابة، جيد أو غير جيد، عالٍ أو متدنٍ...إلخ كما يطلق على مثل هذه المحاولات الثنائية أسم محاولات برنولي (Bernolli trials). والجدير بالذكر هنا: أن التوزيع الثنائي لا يعتمد على محاولة برنولية واحدة فقط بل على عدد من المحاولات التي يجب أن تكون مستقلة عن بعضها البعض لضمان مصداقية الاحتمالات وخاضعة لنفس الشروط والمواصفات بدون تغيير. وفي التوزيع الثنائي

يتم حساب عدد النجاحات (الحصول على المطلوب) التي يرمز لها بالرمز S ولعددها بالرمز x ضمن عدد $n > 0$ من المحاولات البرنولية التي يكون احتمال النجاح فيهـا مسـاوياً P. وعليه يرمز لهذا التوزيع بالرمز $B(n,p)$ وهو من التوزيعات التكراريـة المنقطعـة غيـر المتصلة.

ومن الأمثلة على تطبيقات التوزيع الثنائي هو إنتاج المواد الصناعية التي يمكن أن تكون جيدة أو معطوبة حيث يتم مشاهدة وحساب عدد السلع المعطوبة x ضمن عينة عشـوائية مسـتقلة حيث يأخذ X القيم التالية: $x = 0, 1, 2, \ldots\ldots, n$

ولو افترضنا أنّ احتمال ظهور أو مشاهدة سلعة معطوبة في العينة هو P فإنّ احتمـال عـدم مشاهدة سلعة معطوبة هو $q = 1 - p$ وعليه نجد أنّ احتمال مشاهدة أو ظهور x مـن السـلع المعطوبة في العينة ذات الحجم n من العناصر يكتب بالشكل التالي:

$$\underbrace{ppp\ldots p}_{x\ \text{من المشاهدات}}\ \underbrace{qqq\ldots q}_{n-x} = p^x q^{n-x} = p^x (1-p)^{n-x} \qquad \textbf{(4.1)}$$

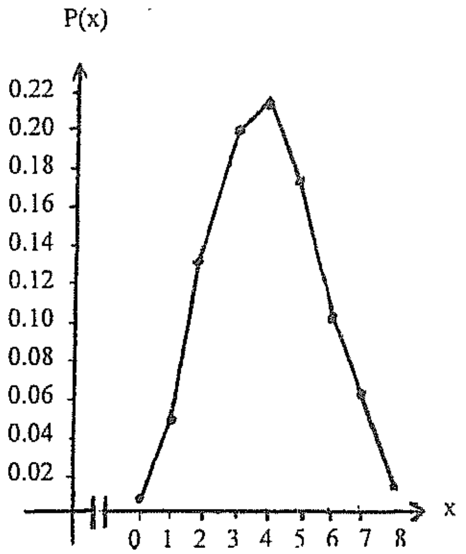
وباستخدام قاعدة التوافيق نجد أنّ تعداد مثل هذا الترتيب أو هذه الحالـة أي ظهـور x مـن النجاحات ضمن n من الحالات يحسب من العلاقة التالية:

$$C_n^x = \frac{n!}{x!(n-x)!} = \binom{n}{x} \qquad \textbf{(4.2)}$$

وعليه تأخذ دالة التوزيع الثنائي الاحتمالية الصورة التالية:

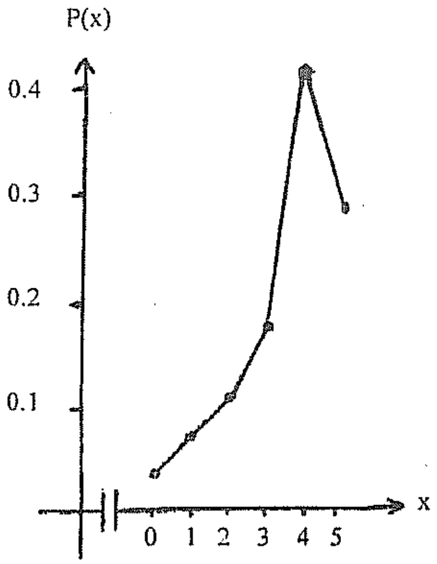
$$f(x,n,p) = P(X = x) = \begin{cases} C_n^x p^x q^{n-x} & x = 0, 1, 2, \ldots\ldots, n \\ 0 & \text{خلاف ذلك} \end{cases} \qquad \textbf{(4.3)}$$

ويمكن تمثيل دالة التوزيع الثنائي بيانياً وفق العلاقـة (4.3) وعلـى افتـراض أنّ $n = 20$ ، $p = 0.20$ بالشكل (4-1) التالي:



الشكل (4-1)

$B(n = 20, P = 0.20)$



الشكل (4-2)

$B(n = 5, P = 0.80)$

بينما يمثل الشكل (4-2) التوزيع الثنائي عندما n=5 ، p=0.80 ويلاحظ من الشكلين تقارب التوزيع الأول من التماثل عندما n كبيرة و p صغيرة وبالعكس في الشكل الثنائي (4-2).

4-1-1 خصائص التوزيع الثنائي:

1- يتقارب التوزيع الثنائي من التوزيع الطبيعي (المتماثل) عندما تتزايد n بشكل كبير جداً. (لاحظ شكل (4.1))

2- $E(X) = nP$ ولتوضيح ذلك نلاحظ أنّ:

$$E(X) = \sum_{x=0}^{n} x.p(x)$$

$$= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} . p^x q^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} . p^x q^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} . p^{x-1} q^{n-x}$$

وبفرض $N = n-1$ و $x-1 = y$ نجد أنّ:

$$E(X) = np \sum_{y=0}^{N} \frac{N!}{y!(N-y)!} p^y q^{n-y}$$

$$= np \sum_{y=0}^{N} C_N^y . p^y q^{n-y}$$

$$= np(p+q)^N = np$$

وذلك لأن: $p + q = 1$

3- $Var(X) = nPq$

ولتوضيح هذه الخاصية أيضاً نجد أنّ:

$$Var(X) = E(X-\mu)^2 = E(X^2) - \mu^2$$

$$E(X^2) = E(X(X-1)+X) = E(X(X-1)) + E(X)$$

$$E(X(X-1)) = \sum_{x=0}^{n} x(x-1)P(x)$$

$$= \sum_{x=0}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)!(n-x)!} p^2 p^{x-2} q^{n-x}$$

$$= n(n-1)P^2 \sum_{y=0}^{m} \frac{m!}{y!(m-y)!} p^y q^{m-y}$$

حيث افترضنا أنّ $y = x-2$ وأنّ $m = n-2$. ومنه:

$$E(X(X-1) = n(n-1)P^2(p+q)^m = n(n-1)P^2$$

ومنه نجد أنّ: $Var(X) = E(X^2) - \mu^2 = n(n-1)p^2 + np - (np)^2 = np(1-p) = npq$

4 - يتقارب التوزيع الثنائي $B(n,p)$ من توزيع بواسون عندما $n \rightarrow \infty$ وأنّ $p = \frac{\lambda}{n}$ حيث λ ترمز لمعلمة (وسيط) بواسون (عدد الأحداث المتوقعة خلال فترة محددة)

مثال (1): إذا كان 40% من طلاب إحدى الكليات لا يملكون سيارات للنقل وإذا كان في إحدى الشعب 8 طلاب فاحسب كلاً من الاحتمالات التالية:

أ- 4 منهم لا يملكون سيارات.

ب- 6 منهم لا يملكون سيارات.

جـ- 2 منهم على الأكثر لا يملكون سيارات.

د- 3 منهم على الأقل لا يملكون سيارات.

الحل: نلاحظ بأن المتغير العشوائي يأخذ صيغة أو مواصفات التوزيع الثنائي لوجود الحدثان المتنافيان يملكون ولا يملكون فقط. ولو افترضنا أن المتغير العشوائي X يمثل عدد الطلاب الذين لا يملكون سيارات في الشعبة المذكورة فإنّ قيم X تكون:

$$x = 0, 1, 2, 3, 4, 5, 6, 7, 8$$

ومن تعريف دالة الكثافة (4.3) نجد أنّ:

$$f(x,8,0.40) = \begin{cases} p(X=x) = \binom{8}{x}(0.40)^x (0.60)^{8-x} & x = 0,1,2,\ldots,8 \\ 0 & \text{خلاف ذلك} \end{cases}$$

وعليه نجد أنّ:

أ- $p(X=4) = f(4,8,0.40) = \binom{8}{4}(0.4)^4(0.6)^4 = 0.2322$

ب- $p(X=6) = f(6,8,0.40) = \binom{8}{6}(0.4)^6(0.6)^2 = 0.0413$

جـ- $p(X \le 2) = p(x=0) + p(x=1) + p(x=2)$

$$= (0.6)^8 + \binom{8}{1}(0.4)^1(0.6)^7 + \binom{8}{2}(0.4)^2(0.6)^6 = 0.315$$

د- $p(X \ge 3) = p(x=3) + \ldots + p(x=8) = 1 - p(x \le 2)$

$$= 1 - 0.315 = 0.685$$

مثال (2): وجد في إنتاج أحد المصانع أن من بين كل 1000 وحدة منتجة توجد 150 وحدة معطوبة فإذا أخذت عينة مكونة من 5 وحدات فاحسب كلاً من الاحتمالات التالية:

أ- الوحدات المختارة كلها سليمة

ب- وحدة واحدة على الأكثر معطوبة

جـ- وحدتان معطوبتان على الأقل

الحل: نلاحظ بأن المتغير العشوائي يتميز بالثنائية (معطوب وغير معطوب) ولو افترضنا بأن X يمثل عدد الوحدات التالفة ضمن العينة العشوائية نجد أنه يأخذ القيم التالية:

$$x = 0, 1, 2, 3, 4, 5$$

كما نجد أنّ: $n = 5$ ، $p = \frac{150}{1000} = 0.15$ $q = 1 - p = 0.85$

وباستخدام التعريف (4.3) السابق نجد أنّ:

أ- الوحدات كلها سليمة يعني $x = 0$ وعليه فإنّ:

$$p(X = 0) = f(0,5,0.15) = \binom{5}{0}(0.15)^0(0.85)^5 = 0.4437$$

ب-

$$p(X \le 1) = p(x = 0) + p(x = 1)$$

$$= 0.4437 + \binom{5}{1}(0.15)^1(0.85)^4 = 0.8352$$

جـ-

$$p(X \ge 2) = 1 - p(x \le 1) - 1 - 0.8352 = 0.1648$$

مثال (3): صندوق يحتوي على عدد من أوراق السحب التي تتضمن خمساً منها فقط على جوائز. فإذا كانت التجارب تشير إلى أنّ نسبة الفوز بجائزة هي 25% فأحسب كلاً من:

أ- دالة كثافة الاحتمال $f(x,n,p)$ للفوز بإحدى الجوائز.

ب- الدالة المولدة للعزوم.

جـ- القيمة المتوقعة والتباين للتوزيع المذكور.

د- $P(0 \le X \le 1)$

هـ- $P(X = 4)$

الحل: أ- من منطوق السؤال نجد أنّ المتغير العشوائي يتميز بالثنائية (الفوز أو عدم الفوز) وبفرض أنّ المتغير X يمثل نتيجة سحب الأوراق فإنّ: $x = 0, 1, 2, 3, 4, 5$ حيث $x = 0$ تمثل سحب ورقة فارغة (بدون جائزة) بينما $x = 1$ تمثل سحب الجائزة الأولى... وهكذا. وعليه فإنّ دالة الكثافة الاحتمالية لهذا التوزيع الثنائي تعطى بالعلاقة التالية:

$$f(x) = \begin{cases} p(X = x) = \binom{5}{x}\left(\frac{1}{4}\right)^x\left(1 - \frac{1}{4}\right)^{5-x} & x = 0, 1, 2, 3, 4, 5 \\ 0 & \text{لغير ذلك} \end{cases}$$

ب-

$$M(t) = \sum_{x=0}^{5} e^{tx} f(x)$$

$$= \sum_{x=0}^{5}\binom{5}{x}\left(\frac{1}{4}e^t\right)^x\left(1 - \frac{1}{4}\right)^{5-x}$$

$$=(\frac{1}{4}e^t+\frac{3}{4})^5=\frac{1}{4^5}(e^t+3)^5$$

جـ- $$\mu=E(X)=M'(0)=\frac{5}{4^5}(e^t+3)^4.e^t\Big|_{t=0}=\frac{5}{4^5}.4^4=\frac{5}{4}$$

$$\sigma^2=M''(0)=\mu^2=np(1-p)=5.\frac{1}{4}.\frac{3}{4}=\frac{15}{16}$$

د- $$P(0\le X\le 1)=\sum_{x=0}^{1}f(x)=\binom{5}{0}(\frac{1}{4})^0(1-\frac{1}{4})^5+\binom{5}{1}(\frac{1}{4})^1(1-\frac{1}{4})^4$$

$$=(\frac{3}{4})^5+\frac{5}{4}(\frac{3}{4})^4=(\frac{3}{4})^4[\frac{3}{4}+\frac{5}{4}]=\frac{81}{128}$$

هـ- $$P(X=4)=f(4)=\binom{5}{4}(\frac{1}{4})^4(1-\frac{1}{4})^{5-4}=5(\frac{1}{4})^4(\frac{3}{4})=15/1024$$

4-2 توزيع بواسون (poisson D.):

يدخل توزيع بواسون في العديد من البحوث التي تشتمل على مشاهدات لأحداث عشوائية تقع أو تتكرر ضمن فترة زمنية محددة أو ضمن حجم معدد أو مسافة أو مساحة محددة مثل تكاثر البكتريا في قطرة خلال فترة زمنية محددة Δt، عدد الاتصالات الهاتفية القادمة أو الخارجة من مدينة أو منطقة أو شركة خلال فترة زمنية محددة، عدد مرات الصيانة لأجهزة مستشفى أو شركة الكترونية...إلخ خلال اليوم الواحد أو الأسبوع الواحد...إلخ.

يتميز توزيع بواسون بحالته الخاصة من التوزيع الثنائي المنقطع ومعلمته λ التي تحسب تكرار وقوع الحدث المستقل خلال فترة زمنية محددة $\Delta t = t_2-t_1 = t-0 = t$ بدأ من الصفر (بداية المشاهدات) كما يتناسب احتمال وقوع أي حدث منفرد مع طول الفترة الزمنية Δt وعلى هذا الأساس يعرف الاحتمال لوقوع أي حدث بحاصل ضرب التكرارات المتوقعة λ (ثابت التناسب) في طول الفترة Δt ولما كانت Δt صغيرة جداً لذا فإنّ هذا التوزيع يتميز بالثنائية أيضاً أي وقوع أو عدم وقوع الحدث خلال الفترة Δt الصغيرة ومنها نجد أنّ المتغير العشوائي X يأخذ القيم التالية 0,1 فقط. وهو ما يفسر عدم استمرارية أو انقطاع التوزيع. وبالاعتماد على المميزات السابقة ودالة التوزيع الثنائي (4.3) نستطيع اشتقاق وتعريف دالة احتمال بواسون والتي تسمى دالة الكثافة الاحتمالية لتوزيع بواسون كما يلي:

$$P(X=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t} \qquad k=0,1,2,\cdots \qquad \textbf{(4.4)}$$

حيث الرمز λ يعني معلمة (وسيط) بواسون والتي تمثل ثابت التناسب خلال الفترة الزمنيـــة t بينما يمثل k عدد تكرارات الحدث خلال الفترة الزمنية المذكورة.

4-2-1 خصائص توزيع بواسون:

1- $Var(X)=E(X)=\lambda t=\mu$

وعندما $t=1$ فإنّ $Var(X)=E(X)=\lambda$

2- $0 \le P(x) \le 1$

3- $\dfrac{P(X=k+1)}{P(X=k)}=\dfrac{\mu}{k+1}$

4- يتقارب التوزيع الثنائي B(n,p) من توزيع بواسون عندما $n \to \infty$ ، $P=\dfrac{\lambda}{n}$

أي أنّ: $\lim_{n\to\infty}\binom{n}{x}p^x q^{n-x}=\frac{\lambda^x}{x!}e^{-\lambda}$

ولتوضيح الخاصية الأولى نجد أنّ: $E(X)=\sum_{k=0}^{\infty}xP(X=k)=\sum_{k=0}^{\infty}k\frac{(\lambda t)^k}{k!}e^{-\lambda t}$

$$=\lambda t\sum_{k=1}^{\infty}\frac{(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t}=\lambda te^{-\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda t)^{k-1}}{(k-1)!}$$

$$=\lambda t\, e^{-\lambda t}.e^{\lambda t}=\lambda t$$

وبالمثل يمكن إثبات أنّ: $Var(X)=\lambda t$

أمّا الخاصية الثانية فتشير إلى أنّ دالة بواسون هي دالة احتمالية لأنّ:

$$\sum_{k=0}^{\infty}P(X=k)=\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}e^{-\lambda t}=e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}$$

$$=e^{-\lambda t}[1+\lambda t+\frac{(\lambda t)^2}{2!}+\dots\dots+\frac{(\lambda t)^k}{k!}+\cdots]$$

$$=e^{-\lambda t}.e^{\lambda t}=1$$

أمّا الخاصية الثالثة فيمكن استقرائها من العلاقة (4.4) كالتالي:

$$P(X=0)=\left[\frac{(\lambda t)^k}{k!}e^{-\lambda t}\right]_{k=0}=e^{-\lambda t}=e^{-\mu}$$

$$P(X=1)=\left[\frac{(\lambda t)^k}{k!}e^{-\lambda t}\right]_{k=1}=\frac{(\lambda t)}{1!}e^{-\lambda t}=\mu e^{-\mu}$$

$$P(X=2)=\frac{\mu}{2}.\mu e^{-\mu}$$

$$P(X=3)=\frac{\mu}{3}.\frac{\mu}{2}.\mu e^{-\mu}$$

ومنها نلاحظ بأن كل حد من حدود بواسون السابقة عبارة عن حاصل الضرب لمتوسط التوزيع μ مقسوماً على رتبة الحد ومضروباً في الحد السابق له أي أنّ:

$$P(X=k+1)=\frac{\mu}{k+1}P(X=k)$$

مثال (4): إذا افترضنا بأنّ المتغير X ذو توزيع بواسون على فترة زمنية $t=4$ ساعة حيث $\lambda=\frac{1}{2}$ هي معلمة بواسون فأوجد قيمة الاحتمال التالي: $P(X\geq 2)$.

الحل: من علاقة التعريف (4.4) والخواص نجد أنّ: $\mu=\lambda t=\left(1/2\right)(4)=2=\sigma^2$

كما نجد أنّ: $k=0,1,2\ldots$ و $P(X=k)=\frac{\mu^k}{k!}e^{-\mu}$

وعليه نجد: $P(X\geq 2)=1-P(X<2)=1-P(X=0)-P(X=1)$

$$=1-\frac{2^0}{0!}e^{-2}-\frac{2^1}{1!}e^{-2}$$
$$=1-0.135-0.270=0.595$$

مثال (5): إذا كان المتغير العشوائي X يمثل عدد المرات التي يتعطل فيها جهاز كهربائي في الأسبوع الواحد يخضع لتوزيع بواسون حيث المعلمة $\lambda=0.4$ فاحسب احتمال أن يعمل الجهاز لمدة ثلاثة أسابيع.

الحل: من العلاقة (4.4) نكتب: $x=0,1,2,\ldots$ $P(X=x)=\frac{\lambda^x}{x!}e^{-\lambda}$

ومنه نجد أنّ احتمال عدم تعطل الجهاز خلال أسبوع واحد هو:

$$P(X=0)=\frac{(0.4)^0}{0!}e^{-0.4}=0.670$$

وأنّ احتمال عدم تعطل الجهاز لمدة ثلاثة أسابيع متواصلة يمكن أن يحسب وفقاً لقاعدة ضـــرب الاحتمالات للأحداث المستقلة كالتالي:

$$P(X=0).P(X=0).P(X=0)=(0.670)^3=0.301$$

مثال (6): يحتوي كتاب 500 صفحة وعلى 300 خطأ موزعة على صفحات الكتاب فأحســب كلاً من الاحتمالات التالية عند اختيار صفحة معينة للقراءة:

أ- لا يوجد أي خطأ في الصفحة المختارة.

ب- يوجد خطأ واحد فقط في الصفحة المختارة.

جـ- وجود خطأين على الأقل في الصفحة المختارة.

الحل: لنفترض أنّ المتغير العشوائي X يمثل عدد الأخطاء في الصفحة الواحدة وهـــو يتبــع التوزيع الثنائي $B(n,p)$ حيث: $x=0,1,2,...,300$

$$P=\frac{1}{500}=0.002 \quad ، \quad n=300$$

وحيث أنّ العدد n كبير نسبياً و p صغيرة جداً لذا يمكن تقريب توزيع ذي الحـــدين الثنـــائي لتوزيع بواسون، وعليه نجد أنّ معلمة بواسون: $\lambda=np=(300).(0.002)=0.6$

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda} \qquad k=0,1,\ldots\ldots300$$

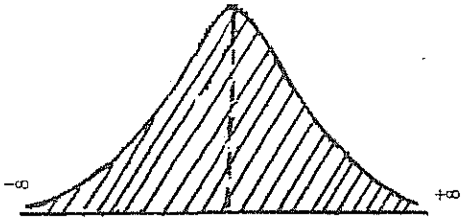
ومنه نجد أنّ: أ- $P(X=0)=\frac{(0.6)^0}{0!}e^{-0.6}=e^{-0.6}$

ب- $P(X=1)=\frac{(0.6)^1}{1!}e^{-0.6}=0.33$

جـ- $P(X\geq 2)=1-P(X=0)-P(X=1)=0.12$

4-3 التوزيع الطبيعي (Normal distribution):

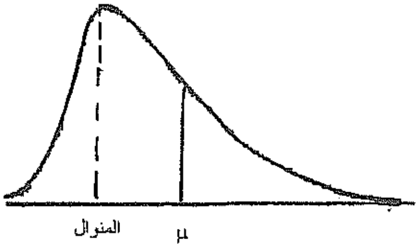
يُقال عن أي توزيع تكراري أو نسبي بأنه توزيع طبيعي ويرمز له بالرمز $N(\mu,\sigma^2)$ حيث μ متوسط التكرارات بينما σ ترمز للانحراف المعياري إذا كانت تكرارات المتغير العشوائي X (كماً كانت أو وصفاً) تنتشر وتتوزع بصورة متماثلة حول متوسط تلك التكرارات وبعبارة أخرى إذا كانت أعلى التكرارات للمتغير X تتمركز وتتقارب من متوسط التوزيع بينما أقلها تنتشر على الأطراف بعيدة عنه وهي التي تمثل القيم أو المواصفات الشاذة في التجربة الإحصائية وبلغة بيانية يُقال عن التوزيع التكراري بأنه طبيعي إذا كان منحني ذلك التوزيع (أو مدرجه التكراري) متماثلاً حول الخط الرأسي المار بمتوسطه μ. وبخلاف ذلك يُقال عن التوزيع بأنه غير طبيعي أو غير متماثل الانتشار أو ملتوٍ (skewed).



-a-

منحنى لتوزيع طبيعي غير ملتوٍ

تنتشر التكرارات بشكل متماثل على طرفي أو جناحي التوزيع



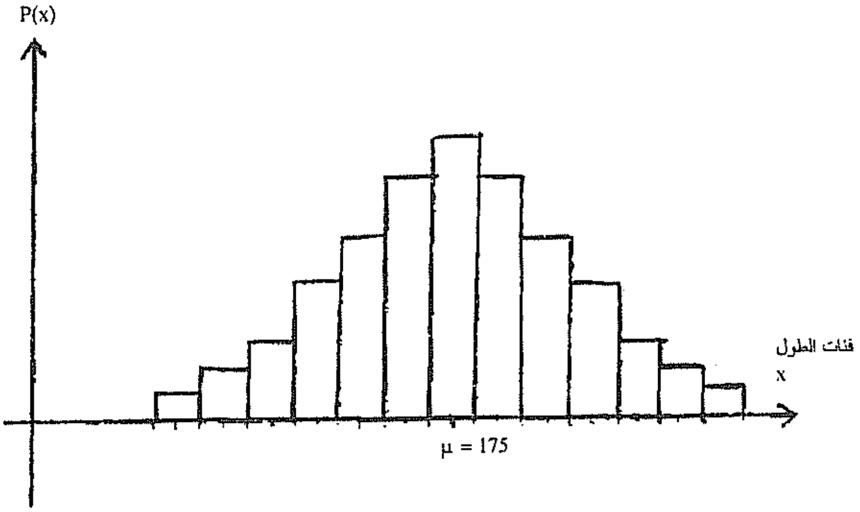
-b-

منحنى لتوزيع ملتو غير طبيعي

تنتشر التكرارات بصورة غير متماثلة على طرفي التوزيع

شكل (4-3)

فعلى سبيل المثال (كما في الأمثلة السابقة الواردة في الفصل الأول) نجد أنّ توزيع أطوال الطلاب الذكور في إحدى جامعات القطر يأخذ شكل المدرج التكراري التالي (شكل 4-4) وهو مدرج طبيعي التوزيع ومتماثل الانتشار حول متوسط الأطوال $\mu=175$. ولو كان تصنيف أو تبويب فئات الأطوال دقيقاً جداً ولنقل مثلاً بالمليمترات أو أجزاء المليمترات لتضائلت أو اختفت المدرجات وتقاربت من المنحني التكراري المتماثل كما في الشكل (4-3-a) أعلاه.



شكل (4-4)

مدرج تكراري يمثل توزيع أطوال الطلبة الذكور في إحدى الجامعات

وبالمثل نجد أن درجات الطلبة تتوزع توزيعاً طبيعياً أيضاً حول متوسطها....إلخ

ويلاحظ بأن مجموع مساحات المستطيلات يساوي مجموع النسب أو الاحتمالات وكلما تضائلت المدرجات واختفت الخطوط المنكسرة أي كلما كان طول فترات التقسيم صغيراً أي $\Delta x \to 0$ كلما تقارب المدرج التكراري من المنحني وعندئذٍ نجد أنّ:

المساحة تحت المنحني من a إلى b = $P(a \le X \le b) = \lim_{\Delta x \to 0} \sum_{x=a}^{b} P(x)\Delta x = \int_a^b f(x)dx$

ويلاحظ من التجارب المتكررة بأن التوزيع الطبيعي يتمثل بمنحني أملس متصل يأخذ شكل الجرس المقلوب ويعرّف رياضياً بدالة الكثافة الاحتمالية لمتغيره العشوائي المستمر X التالية:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (4.5)$$

حيث نطاق الدالة $D_f = \{x : -\infty < x < \infty\}$

وبحساب التوقع من العلاقة السابقة نجد أنّ:

$$E(X) = \int_{-\infty}^{+\infty} x f(x)dx = \mu \qquad (4.6)$$

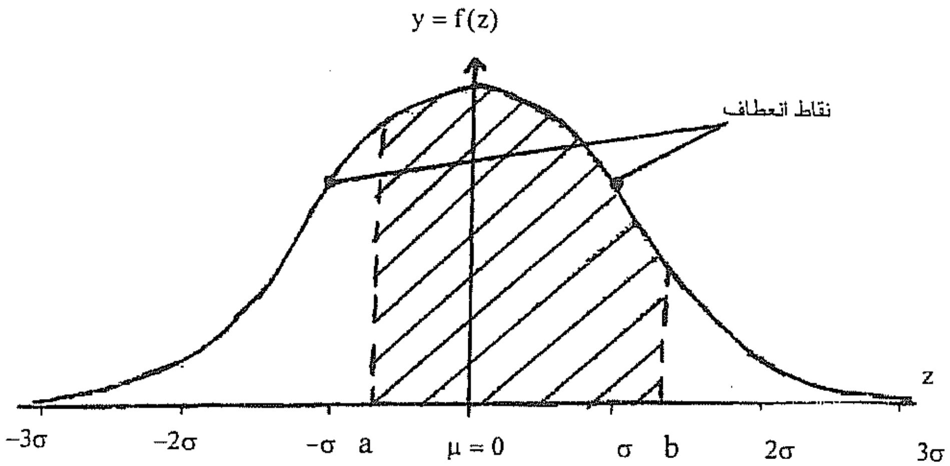
كما نجد أنّ التباين لهذا المتغير يكون:

$$\sigma^2 = Var(X) = E\left((X-\mu)^2\right) = \int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2 \qquad \textbf{(4.7)}$$

ومن أمثلة التوزيعات الطبيعية الشهيرة هو ما يسمى بالتوزيع الطبيعي المعياري أو القياسي والذي يرمز له بالرمز $N(\mu,\sigma^2) = N(0,1)$ حيث متوسطه صفراً وانحرافه المعياري الواحد وفيه يتم اعتماد الدرجة المعيارية Z بدلاً من المتغير العشوائي X وعليه تكون دالة الكثافة لهذا التوزيع هي:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \textbf{(4.8)}$$

والتي تأخذ بيانياً الشكل التالي:



الشكل (4-5)

المنحني الطبيعي المعياري الذي يمثل التوزيع الطبيعي المعياري $N(0,1)$

وهو التوزيع المستخدم بكثرة نظراً لتطبيقاته الواسعة وحساباته المجدولة [ملحق رقم I] والذي تمثل المساحات تحت المنحني فيه احتمال وقوع المتغير العشوائي Z ضمن النطاق المحدد. أي أنّ:

$$P(a \le Z \le b) = \int_a^b f(z)dz \qquad \textbf{(4.9)}$$

4-3-1 خواص التوزيع الطبيعي $N(\mu,\sigma^2)$:

يمكننا تلخيص خواص التوزيع بالنقاط التالية:

1- قيم المتوسط الحسابي μ والانحراف المعياري σ تحددان كمية ونوعية الانتشار والتشتت للبيانات والمشاهدات للمتغير العشوائي X في التجربة الإحصائية كما في الشكلين التاليين (4-6 a,b) حيث نجد منهما أنّ:

68.27٪ من البيانات أو قيم المتغير العشوائي تقع ضمن الفترة $\mu - \sigma \le X \le \mu + \sigma$

95.45٪ من البيانات أو قيم المتغير العشوائي تقع ضمن الفترة $\mu - 2\sigma \le X \le \mu + 2\sigma$

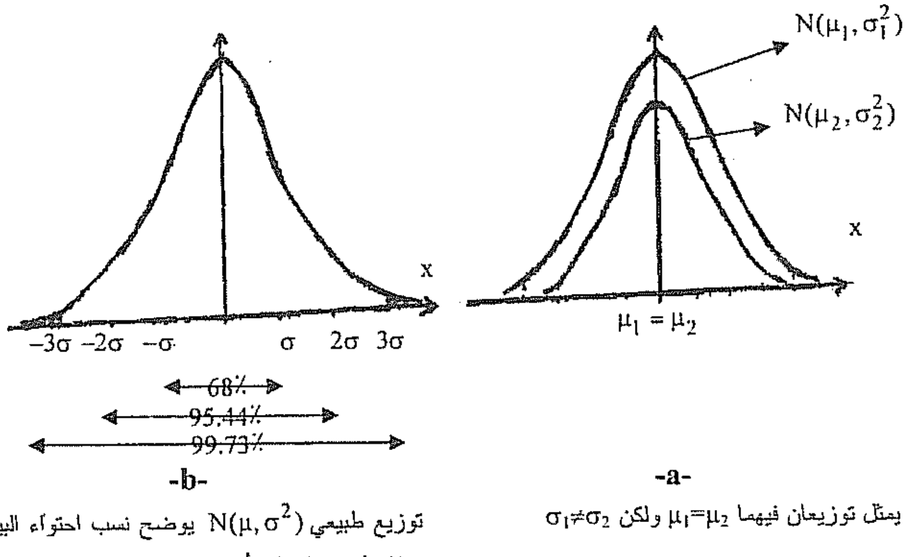
99.73٪ من البيانات أو قيم المتغير العشوائي تقع ضمن الفترة $\mu - 3\sigma \le X \le \mu + 3\sigma$

حيث يعبر عن نسب هذه البيانات بشكل احتمالات كما يلي وعلى الترتيب:

$$P(\mu - \sigma \le X \le \mu + \sigma) = 0.6827$$

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 0.9545$$

$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0.9973$$



-a-

يمثل توزيعان فيهما $\mu_1=\mu_2$ ولكن $\sigma_1 \ne \sigma_2$

-b-

توزيع طبيعي $N(\mu,\sigma^2)$ يوضح نسب احتواء البيانات ويلاحظ من الشكل بأن μ تحدد مركز التوزيع بينما يحدد الانحراف المعياري انتشار وارتفاع التوزيع.

الشكل (4-6)

2- المساحة الكلية تحت المنحني الطبيعي تساوي 1 ولذلك تُعتمد دالة الكثافة كدالة احتمالية.

3- يُشكل المحور الأفقي أي محور المتغير العشوائي X خطاً مقارباً أفقياً للمنحني الطبيعي.

4- يكون المنحني الطبيعي (التوزيع الطبيعي) معيارياً (أو قياسياً) إذا كان متوسطه $\mu=0$ وانحرافه المعياري $\sigma=1$.

ملاحظة ①: لاستخدام الجداول الخاصة بالتوزيع الطبيعي المعياري تحسب قيمة الاحتمال من البيانات في العمود الأول تحت الدرجة المعيارية z أو من البيانات في الأعمدة الأخرى إن كانت z ذات قيم عشرية أخرى. فمثلاً إذا كانت $z=2.15$ فإننا نقرأ قيمة الاحتمال في العمود 0.05 والسطر الموافق للرقم 2.1 في العمود z. وهكذا...

ملاحظة ②: بعد معرفة العلاقة بين التوزيع الطبيعي $N(\mu,\sigma^2)$ والتوزيع الطبيعي المعياري $N(0,1)$ يمكننا إجراء التحويلات من الدرجات الطبيعية إلى المعيارية وبالعكس ولاستخدام الجداول الخاصة بالتوزيع الطبيعي المعياري لحساب الاحتمالات الطبيعية أو المقاربة لها يتوجب تحويل البيانات (x) إلى الدرجات الطبيعية المعيارية z أولاً ومن ثم استخدام الملاحظة ① فمثلاً في مجتمع طبيعي التوزيع $N(\mu,\sigma^2)$ غير معياري نجد:

$$\begin{aligned} P(a \le X \le b) &= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) \\ &= P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right) \\ &= \int_{-\infty}^{\frac{b-\mu}{\sigma}} f(z)dz - \int_{-\infty}^{\frac{a-\mu}{\sigma}} f(z)dz \end{aligned}$$

حيث تحسب قيم التكاملات الأخيرة من الجداول الخاصة بالتوزيع المعياري وفقاً للملاحظة ①.

مثال (7): أوجد كلاً من الاحتمالات التالية لتجربة إحصائية عشوائية يتوزع متغيرها العشوائي توزيعاً طبيعياً معيارياً.

أ- $P(Z \le 1.72)$ ب- $P(Z \le -0.54)$

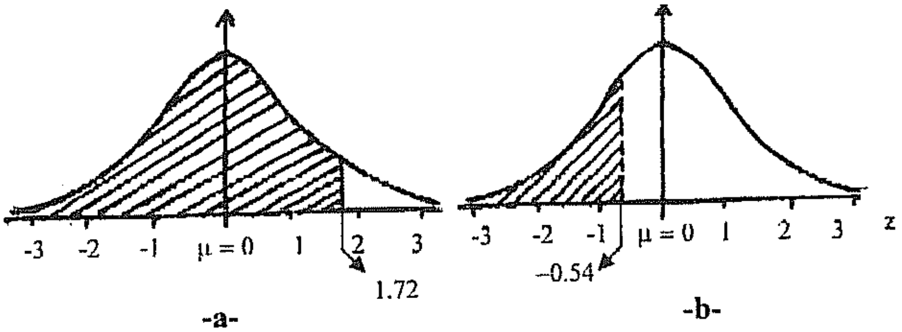
جـ- $P(Z \ge 0.29)$ د- $P(-1.91 \le Z \le 1.45)$

الحل: بما أنّ الدرجات طبيعية معيارية لذا نستخدم الجداول مباشرة لحساب الاحتمالات كالتالي:

أ-

$$P(Z \le 1.72) = F(z) = \int_{-\infty}^{1.72} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

ومن الجدول الخاص نجد أمام الدرجة $z = 1.7$ وفي العمود 0.02 قيمة الاحتمـال وتسـاوي 0.9573 والتي تمثل المساحة تحت المنحني ولغاية النقطة $z = 1.72$



الشكل (4-7)

ب-

$$P(Z < -0.54) = F(z) = \int_{-\infty}^{0.54} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 0.2946$$

جـ -

$$P(Z \ge 0.29) = \int_{0.29}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

$$= 1 - P(Z \le 0.29) = 1 - \int_{-\infty}^{0.29} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

$$= 1 - 0.6141 = 0.3859$$

د -

$$P(-1.91 \le Z \le 1.42) = P(Z \le 1.42) - P(Z \le -1.91)$$

$$= \int_{-\infty}^{1.42} f(z)dz - \int_{-\infty}^{-1.91} f(z)\,dz$$

$$= 0.92220 - 0.02807 = 0.89413$$

مثال (8): وجد في تجربة إحصائية أن فترة العمل والضمانة لأحدى محركات الديزل والتي تعتبر متغيراً عشوائياً طبيعياً ذات متوسط 266 أسبوع وبانحراف معياري $\sigma=12$ أسبوع فما هو احتمال أو نسبة آلات الديزل التي تعمل بين 260 إلى 270 أسبوعاً.

الحل: نفترض أنّ X يمثل فترة العمل لهذا النوع من المحركات الخاضعة للتوزيع $N(266,144)$ والمطلوب هو حساب احتمال أن تكون هذه الفترة أكبر من 260 أسبوعاً وأصغر من 270 أسبوعاً فقط أي: $P(260 \leq X \leq 270)$.

وبما أن المتغير X طبيعي وغير معياري لذا يتوجب تحويل القيم إلى الدرجات المعيارية أولاً ليتسنى لنا استخدام الجداول المعيارية، كالتالي:

$$\begin{aligned} P(260 \leq X \leq 270) &= P\left(\frac{260-266}{12} \leq Z \leq \frac{270-266}{12}\right) \\ &= P(-0.50 \leq Z \leq 0.33) \\ &= P(Z \leq 0.33) - P(Z \leq -0.50) \\ &= 0.6293 - 0.3085 = 0.3208 = 32.08\,\% \end{aligned}$$

4-3-2 التقارب مع التوزيع الطبيعي:

تتقارب الكثير من توزيعات الوسطاء الإحصائية مع التوزيع الطبيعي مثل التوزيع الثنائي $B(n,p)$ عندما تكون n كبيرة جداً و p صغيرة وعموماً يتقارب توزيع الوسيط الإحصائي $\overline{X}$ من التوزيع الطبيعي سواءً كان المجتمع الإحصائي الأم للعينات طبيعي التوزيع أم غير طبيعي التوزيع ففي الحالة الأولى يكون من البديهي استنتاج طبيعية توزيع $\overline{X}$ أمّا في الحالة الثانية أي عندما يكون توزيع المجتمع الأم غير طبيعي فإن نظرية النهاية المركزية تشير إلى تعميم طبيعية توزيع $\overline{X}$ والتي يمكن إثباتها بإجراء التجارب المتكررة على العينات كبيرة الحجم. فإذا كان لدينا مجتمع ما متوسطه μ وانحرافه المعياري σ وأخذنا منه عينات مختلفة لها نفس الحجم n ثم حسبنا $\overline{x}$ لكل عينة فنجد أنّ قيم $\overline{X}$ تتغير من عينة لأخرى ولكنها تبقى قريبة من μ، وبالتالي يمكن اعتبار $\overline{X}$ متغيراً عشوائياً ينتشر بشكل طبيعي حول المتوسط μ وهو ما تنص عليه النظرية التالية:

نظرية ⓫ (نظرية النهاية المركزية):

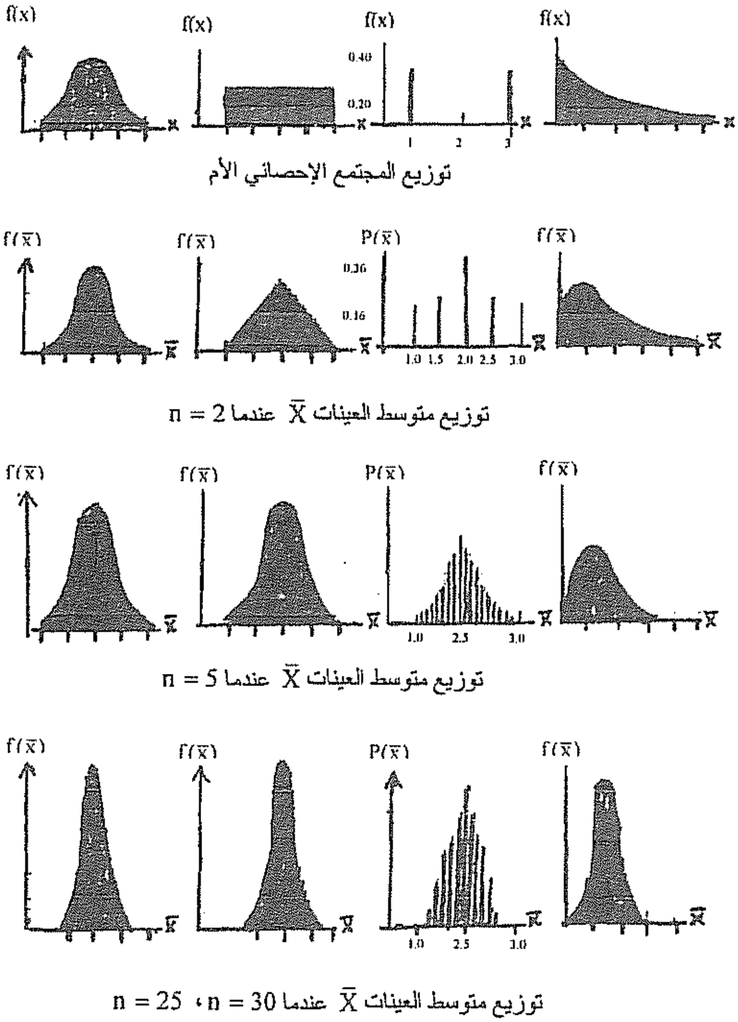
إذا أخذنا عينات عشوائية بأحجام كبيرة بشكل كاف (حجم كل منها n حيث: $n \to \infty$) من أي مجتمع إحصائي (حتى ولو كان غير طبيعي) متوسطه μ وانحرافه المعياري σ، فإنّ المتغير العشوائي $\overline{X}$ يتوزع توزيعاً مقارباً لتوزيع طبيعي متوسطه $\mu_{\overline{x}} = \mu$ وانحرافه المعياري $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$ ويكون للمتغير: $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ التوزيع الطبيعي المعياري.

ملاحظة ③: يكون التقريب في النظرية السابقة جيداً إذا كان $n \geq 30$ بغض النظر عن شكل المجتمع الأم. أمّا إذا كان المجتمع الأم طبيعياً فإنّ $\overline{X}$ يتوزع توزعاً طبيعياً مهما تكن n.

نستعرض فيما يلي بعض نتائج التجارب البيانية لعينات مأخوذة بأحجام مختلفة من مجتمعات إحصائية أصلية مختلفة التوزيع: طبيعي (Normal)، متجانس (Uniform)، ثنائي (Binomial) وأسي (Exponential) على الترتيب أنظر الشكل (4-8) ورصدت نتائج توزيع المتوسط $\overline{X}$ للحالات على الترتيب في الجدول التالي:

الحالة	حجم العينة	التوزيع الأم طبيعي	التوزيع الأم متجانس	التوزيع الأم ثنائي	التوزيع الأم أسي
أ	$n = 2$	طبيعي	مدرج متماثل	مدرج متماثل	شبه متماثل أو منحرف
ب	$n = 5$	طبيعي	مضلع متماثل	مدرج طبيعي	شبه متماثل أو منحرف
جـ	$n = 30$	طبيعي	مضلع طبيعي	مدرج طبيعي	طبيعي

ويلاحظ من الجدول السابق بأنه على الرغم من اختلاف توزيعات المجتمعات الإحصائية الأم إلاّ أنّ توزيع $\overline{X}$ يأخذ في النهاية الشكل الطبيعي عندما يكون حجم العينة كبيراً وكافياً ولو ازداد حجم العينة فإنّ توزيع $\overline{X}$ سيتقارب مع التوزيع الطبيعي بدقة أكبر خصوصاً إذا كان $n > 30$.



الشكل (4-8): يوضح نماذج بيانية لتجارب المعاينة على مجتمعات أصلية متنوعة التوزيع ويتقارب توزيع $\overline{X}$ مع الطبيعي كلما ازداد حجم العينة

مثال (9): أوجد قيمة $\overline{X}_1$ لتوزيع المتوسطات لـ $\overline{X}$ والتي تقع على يسارها 25% من قيم $\overline{X}$ إذا اخترنا عينة عشوائية حجمها $n=100$ من مجتمع إحصائي متوسطه $\mu = 40$ وتباينه $\sigma^2=4$.

الحل: بما أنّ n>30 فإنّ توزيع المتغير $\overline{X}$ يكون طبيعياً بمتوسط $\mu_{\bar{x}} = \mu = 40$ وتباين $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{4}{100}$ وذلك حسب نظرية النهاية المركزية. وبالتحويل إلى التوزيع الطبيعي المعياري نجد أنّ:

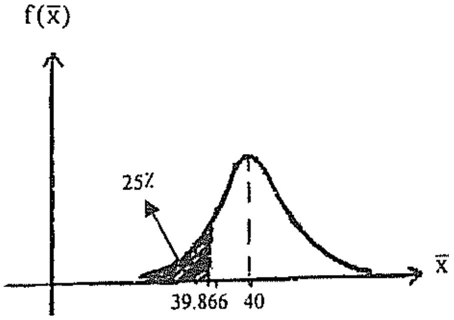
$$z = \frac{\bar{x}_1 - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x}_1 - 40}{2/10}$$

ومنه:
$$\bar{x}_1 = (0.2)z + 40$$

وبما أنّ لدينا:
$$P(\overline{X} < \bar{x}_1) = 0.25 \Leftrightarrow P(Z < z) = 0.25$$

حيث نجد من جدول التوزيع الطبيعي المعياري أنّ: $Z = -0.67$

وبالتالي:
$$\bar{x}_1 = (0.2)(-0.67) + 40 = 39.866$$



مثال (10): لنفترض أنّ المتغير X يمثل متوسط درجات طالب أُختير عشوائياً من معهد خاص فإذا كان لتوزيع X متوسط قدره $\mu=2.5$ وانحراف معياري قدره $\sigma=0.4$ وأخذت عينة عشوائية تتكون من 36 طالباً وحسب متوسط متوسطات الطلبة $\bar{x}$ فما هو احتمال أن يقل المتغير العشوائي $\overline{X}$ عن 2.4 وما هو احتمال أن يقع $\overline{X}$ ضمن الفترة الاحتمالية $2.4 \le \overline{X} \le 2.7$

الحل: بالرغم من أنّ توزيع X هو غير طبيعي ولكن حجم العينة n=36 يبرر استخدام النظرية ❶ وافتراض أن توزيع $\overline{X}$ سيكون طبيعياً وعليه فإنّ:

انحراف توزيع المتوسطات:
$$\sigma_{\bar{x}} = \frac{0.4}{\sqrt{36}} = 0.067$$

والدرجة المعيارية المقابلة للقيمة $\bar{x} = 2.4$ تكون:

$$z=\frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}=\frac{2.4-2.5}{0.067}=-1.50$$

حيث نجد أنّ: $P(\overline{X}<2.4)=P(Z<-1.50)=0.07$

كما نجد أنّ احتمال وقوع المتوسط ضمن الفترة [2.4,2.7] هو:

$$P(2.4\le\overline{X}\le 2.7)=P\left(\frac{2.4-2.5}{0.067}<Z<\frac{2.7-2.5}{0.067}\right)$$
$$=P(-1.5<Z<3)=P(Z<3)-P(Z<-1.50)$$
$$=0.9987-0.0668=0.9319$$

عند تطبيق نظرية النهاية المركزية نحتاج لمتوسط وتباين المجتمع الأم μ و σ^2، ولكن في كثير من الحالات لا يُستطاع حساب تباين المجتمع σ^2 ولذا نقبل، من أجل عينات حجمها $n\ge 30$ ، تباين العينة S^2 كتقدير جيد للتباين σ^2 حيث لا تتغير S^2 كثيراً من عينة لأخرى وبالتالي فإنّ نظرية النهاية المركزية تبقى محققة هنا وتصبح الدرجة المعيارية z بالشكل: $z=\frac{\bar{x}-\mu}{S/\sqrt{n}}$ أمّا إذا كان حجم العينة n<30 فإنّ قيم S^2 تتغير من عينة لأخرى بشكل واضح وقيم $z=\frac{\bar{x}-\mu}{S/\sqrt{n}}$ في هذه الحالة لا تتبع التوزيع الطبيعي المعياري. لذا سنستخدم توزيعاً جديداً يدعى بتوزيع ستيودنت أو توزيع **t**.

4-4 توزيع (ستيودنت) t: Student T-distribution

يرجع توزيع **t** إلى الباحث غوست (W.S Gosset) الذي عرض أبحاثه تحت أسم ستيودنت ويعتبر توزيع T واحداً من التوزيعات الطبيعية المتماثلة حول متوسطها وفقاً للتجارب التي أجراها غوست على مجتمعات طبيعية التوزيع وغير معروفة التباين σ^2 وأثبت أن هذه الحالة أي أنّ معرفة التباين لمجتمع طبيعي غير ضرورية نظراً لإمكانية استبدال تباين العينات العشوائية S^2 كقيمة بديلة لتباين المجتمع الأم لأنّ: $E(S^2)=\sigma^2$

وتتلخص تجارب ستيودنت على توزيع متوسطات عينات عشوائية مأخوذة من مجتمع إحصائي (لانهائي) طبيعي التوزيع. فوجد أنّ هذه المتوسطات تتوزع توزيعاً طبيعياً وعليه عرّف درجته المعيارية ذات درجة الحرية $v=n-1$ لهذا التوزيع بالنظرية التالية:

نظرية ④ (نظرية ستيودنت):

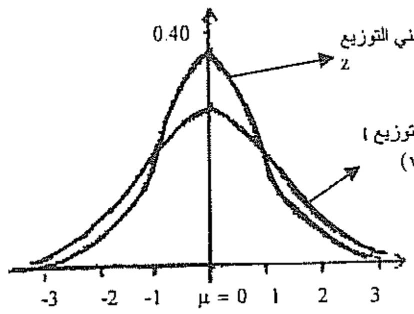
إذا كان كل من $\overline{X}$ ، S^2 هما المتوسط والتباين لعينة عشوائية مأخوذة من مجتمع إحصائي طبيعي التوزيع متوسطه μ فإنّ المتغير العشوائي:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \qquad (4.10)$$

يتوزع توزيعاً طبيعياً أيضاً وبدرجة حرية $v = n - 1$.

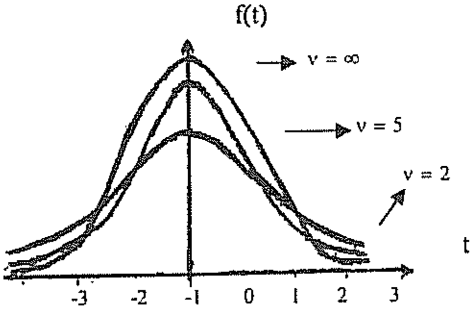
ويشابه هذا التوزيع الجديد التوزيع الطبيعي المعياري من حيث كون متوسطه $\mu = 0$ إلاّ أنّ تباينه $\neq 1$

حيث أنّ تغير قيم S من عينة لعينة أخرى وعدم ثباته يسبب تفلطحاً أكبر وانتشاراً أبعد للتوزيع كما في الأشكال التالية:



شكل (4-9)

يوضح الفرق بين منحني التوزيع المعياري z والتوزيع t عندما يكون حجم العينة n = 4 في إحدى التجارب الإحصائية.



شكل (4-10)

منحني توزيع t وبدرجات حرية مختلفة $v=2,5,\infty$.
يلاحظ فيه تفلطح المنحني، وتدببه كلما قلت أو زادت درجات الحرية (حجم العينة) على الترتيب كما يلاحظ تقارب قيم t من بعضها للعينات الكبيرة.

والجدير بالذكر هنا هو أنّ توزيع t يعتبر مناسباً جداً للتحليل الإحصائي عندما يكون المجتمع الأم طبيعي التوزيع (وغير معروف التباين σ^2) على الرغم من استخدامه في خلاف ذلك أحياناً في التوزيعات المقاربة للتوزيعات الطبيعية كما أنه يستخدم وبصورة فعّالة ليس فقط للعينات صغيرة الحجم $n<30$ بل حتى للعينات كبيرة الحجم أيضاً على الرغم من أنّ جداوله

الحسابية لا تتعدى $v = 30$ وذلك لأن الاحتمالات عندئذٍ تكون متقاربة جداً من بعضها البعض كما يوضحه الشكل السابق (4-10).

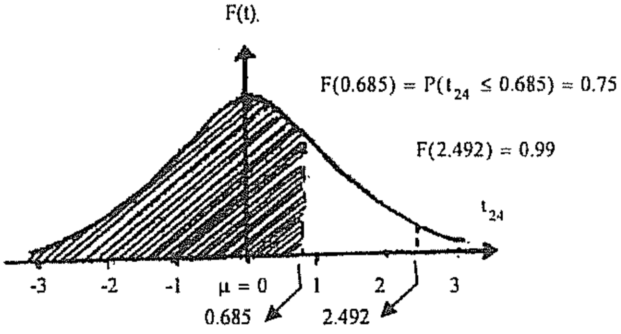
من التعريف السابق للدرجة المعيارية t نجد أنها تمثل دالة بالمتغير $v = n - 1$ وعلى هذا الأساس حسبت قيم t وفقاً لدرجات الحرية والتي تمثل عدد عناصر العينة العشوائية أيضاً ونظمت جداول قيم t الخاصة والتي يمكن مراجعتها بقراءة الرقم مقابل الدرجة v والتي تمثل بيانياً المساحة تحت المنحني عند الدرجة المذكورة كما تمثل إحصائياً احتمالية حصول أو وقوع الحدث أو القيمة x وفقاً للعمود الذي يبيّن ذلك الاحتمال. حيث نجد مثلاً عندما n=25 أنّ $v = 24$ وأنّ: $P(t_{24} < 2.492) = F(t_{24}) = 0.99$ حيث 0.99 يمثل العمود الخامس في جدول t. أمّا F(t) وهي دالة الاحتمال التراكمية (المجمّعة) والمحسوبة في الجداول فهي معرفة بالتكامل المعتل التالي:

$$F(t) = \int_{-\infty}^{t} \frac{(\frac{v-1}{2})!}{(\frac{v-2}{2})!\sqrt{\pi n}(1+\frac{t^2}{v})^{\frac{v+1}{2}}} dt \qquad (4.11)$$

حيث تتوضح العلاقة مع دالة غاما $\left(\Gamma(\frac{v+1}{2}) = (\frac{v-1}{2})!\right)$

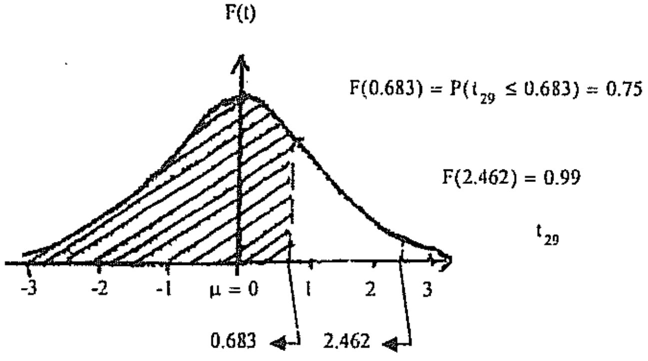
وفيما يلي بعض الأشكال التي توضح حسابات الاحتمال التراكمي F(t) وعلاقة أحجام العينات أو درجات الحرية

عندما يكون حجم العينة n=25 فإنّ v=n-1=24 وإنّ $F(t_{24})$ يأخذ الشكل التالي:



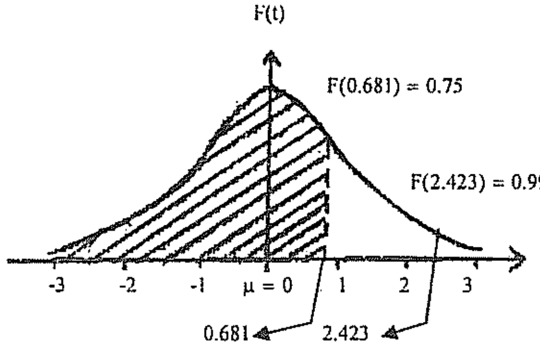
شكل (a)

وعندما يكون حجم العينة n=30 فإنّ $v = 29$ وإنّ $F(t_{29})$ يأخذ الشكل التالي:



شكل (b)

وعندما يكون حجم العينة n=41 فإنّ v=40 وإنّ $F(t_{40})$ يأخذ الشكل التالي:



$$t_{40} = \frac{\bar{x}-\mu}{S_{\bar{x}}/\sqrt{41}}$$

شكل (c)

الشكل (4-11) لاحظ أنّ قيم t_{24} , t_{29} , t_{40} تكاد تكون متساوية في التوزيعات الثلاثة ولنفس الاحتمالات (المساحات) المأخوذة في الأشكال الثلاث.

مثـــال (11): تدعي شركة صناعية للمواد البلاستيكية أنّ متوسط إنتاجها الأسبوعي من الأطقم البلاستيكية هو: μ=825 طقماً وأنّ التباين غير معلوم. وللتأكد من هذا الادعاء، اختيرت عينة عشوائية ذات حجم n=25 من منتجات هذه الشـــركة، فكـــان متوســـط الإنتـــاج الأســـبوعي $\bar{x} = 780$ طقماً وبانحراف معياري S=105. فإذا علمت أنّ منتجات الشركة تتوزع توزيعـــاً

طبيعياً على وجه التقريب أوجد احتمال أن يكون متوسط الإنتاج الأسبوعي للشركة 780 على الأكثر.

الحل: لحل مثل هذه المسائل يتوجب تحويل المشاهدات أو البيانات إلى الدرجة التائية وبدرجة حرية $\nu = n - 1 = 24$ وذلك كما يلي:

$$P(\overline{X} \le 780) = P\left(t_{24} \le \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}\right) = P\left(t_{24} \le \frac{780 - 825}{\frac{105}{\sqrt{25}}}\right)$$

$$= P(t_{24} \le -2.143) = 1 - P(t_{24} \ge 2.143)$$

وبما أن الدرجة المعيارية التائية 2.143 لا تظهر من جدول t_{24} لذا يمكننا تقديرها بالطريقة الفتروية التالية حيث نجد من الشكل (4-12) أنّ:

$$P(t_{24} \le 2.064) = 0.975 \qquad P(t_{24} \le 2.492) = 0.990$$

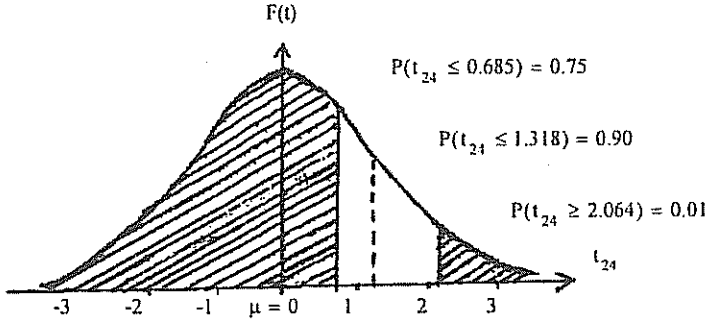
وهذا يعني أنّ: $P(t_{24} \le 2.064) \le P(t_{24} \le 2.143) \le P(t_{24} \le 2.492)$

أو أنّ:

$$1 - P(t_{24} \le 2.492) \le P(\overline{X} \le 780) = P(t_{24} \le -2.143) \le 1 - P(t_{24} \le 2.064)$$

$$0.01 \le P(\overline{X} \le 780) \le 0.025$$

وهذا يعني أنّ الحصول على 780 طقماً في الأسبوع على الأكثر يحدث بنسبة 1٪ إلى 2.5٪.



الشكل (4-12): بعض القيم الجدولية لتوزيع ستيودنت t_{24} عندما n=25

4-4-1 خصائص التوزيع t:

1- القيمة المتوقعة للدرجة المعيارية التائية صفرية أيّ: $E[t] = 0$

2- التباين لتوزيع t عندما $n \geq 3$: $Var(t) = \frac{\nu}{\nu-2}$

3- يستخدم توزيع t بصورة خاصة عند عدم معلوميه تباين المجتمع الإحصائي الأم.

4- تتقارب قيم t من بعضها البعض عندما تكون أحجام العينات كبيرة.

5- يعتمد توزيع t على شرط التوزيع الطبيعي للمجتمع الإحصائي الأم وبخلاف ذلك قد لا تكون نتائجه دقيقة.

مثال (12): في تجربة لقياس الأقطار الداخلية لمكابس المحركات والتي تتوزع توزيعاً طبيعياً حول متوسطها $\mu=120$ ملمتر. تم أخذ عينة عشوائية حجمها $n=5$ وحسبت الأقطار التالية لها: $x_1 = 121 \quad x_2 = 120.5 \quad x_3 = 122 \quad x_4 = 119 \quad x_5 = 119.5$

أحسب احتمال أن تكون الأقطار ضمن الفترة $119.5 \leq \overline{X} < 120.5$

وأحسب احتمال أن تكون الأقطار أكبر من 121 ملم أو أصغر من 118 ملمتر

الحل: بما أن المجتمع الأم طبيعي التوزيع وغير معروف التباين σ^2 لذا نستخدم توزيع t لحساب الاحتمالات المذكورة حيث نجد أنّ:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{602}{5} = 120.4$$

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}} = \sqrt{\frac{(0.6)^2+(0.1)^2+(1.6)^2+(1.4)^2+(0.9)^2}{5}}$$

$$= \sqrt{\frac{0.36+0.01+2.56+1.96+0.81}{5}} = \sqrt{\frac{5.70}{5}} = \sqrt{1.14} = 1.068$$

وبما أنّ الدرجة المعيارية التائية: $t = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ وبتحويل الفترات المطلوبة إلى الفترات التائية نستطيع استخدام الجداول الخاصة بتوزيع **t** كالتالي:

$$P(119.5 \leq \overline{X} \leq 120.5) = P\left(\frac{119.5-120}{0.478} \leq \frac{\overline{x}-\mu}{S/\sqrt{n}} \leq \frac{120.5-120}{0.478}\right)$$

$$= P(-1.046 \leq t_4 \leq +1.046) = P(t_4 \leq 1.046) - P(t_4 < -1.046)$$

ونظراً لتماثل منحني t نجد أنّ:

$$= P(t_4 \le 1.046) - [1 - P(t_4 \le 1.046)]$$
$$= 2P(t_4 \le 1.046) - 1$$

وبما أنّ الدرجة 1.046 ليست موجودة في ا لجدول الخاص لذا نتوقع أن يكون احتمالها بين (0.75) و (0.90) ولنقل مثلاً أنّ الاحتمال هو 0.84 وعليه نجد أنّ الاحتمال:

$$P(119.5 \le \overline{X} \le 120.5) = 2(0.84) - 1 = 0.68$$

وبنفس الأسلوب يمكن حساب:

$$P(\overline{X} \ge 121) = P(t_4 \ge \frac{121-120}{1.068/\sqrt{5}}) = P(t_4 \ge \frac{1}{1.046}) = P(t_4 \ge 0.956)$$
$$= 1 - P(t_4 \le 0.956) \cong 1 - 0.85 = 0.15$$
$$P(\overline{X} \le 118) = P(t_4 \le \frac{118-120}{1.068/\sqrt{5}}) = P(t_4 \le -1.912)$$

ونظراً لتماثل المنحني نجد أنّ:

$$= 1 - P(t_4 \le 1.912) \cong$$
$$= 1 - 0.93 = 0.07$$

ويلاحظ بالمقارنة أنّ احتمال أن يكون القطر أكبر من 121 أعلى من احتمال أن تكون الأقطار صغيرة وأقل من 118. ولسنا هنا بشأن اتخاذ قرار حول الإنتاج أو الجودة لأن له مجال آخر في الفصول القادمة.

4-5 توزيع مربع كاي (χ^2): chi-square

يعتبر التوزيع التكراري الاحتمالي لمربع كاي من التوزيعات التكرارية المتصلة والذي يستخدم بكثافة للإجابة عن الكثير من التساؤلات حول المجتمع الإحصائي الام مثل التوافق بين مجاميع معينة من المعلومات (Goodness of fit) أو التجانس بين مجموعة من العينات أي ما إذا كانت مأخوذة أو تابعة لنفس المجتمع الإحصائي.

وتعرف دالة الكثافة الاحتمالية لهذا التوزيع للأعداد الحقيقية غير السالبة فقط لأنّ $\chi^2 \ge 0$ وتأخذ الشكل التالي:

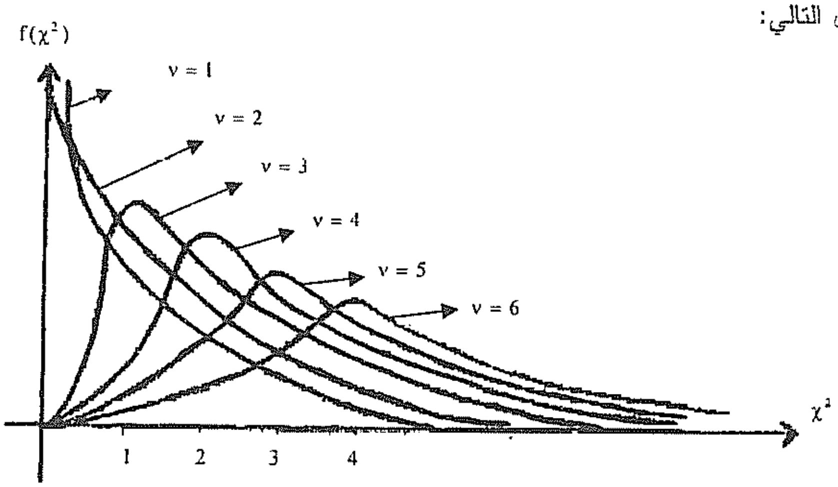
$$f(\chi^2_{(\alpha,\nu)}) = P(\chi^2_{(\alpha,\nu)})d\chi^2 = \frac{e^{-\frac{1}{2}\chi^2}.(\chi^2)^{\frac{1}{2}\nu-1}}{2^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu)} \qquad \textbf{(4.12)}$$

حيث الرمز χ^2 والذي يسمى إحصائية بيرسون أحياناً يعرّف بالعلاقة التالية:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \qquad \textbf{(4.13)}$$

وحيث k = عدد المعلومات، $\nu = n - 1$ ، Γ ترمز لدالة غاما الخاصة بينما e_i تمثل القيمة المتوقعة للمعلومة الميدانية أي الاحتمال النسبي للمتغير العشوائي Y ، O_i تمثل القيمة الفعلية الميدانية للمتغير العشوائي Y نفسه $n = \sum_{i=1}^{k} O_i$ ، $e_i = nP_i$ ، α احتمال الدرجة الحرجة.

ويمكن تمثيل الدالة (4.12) بيانياً بمنحني متصل وعلى اختلاف درجات الحرية ν كما في الشكل التالي:



الشكل (4-13)

منحنيات دالة الكثافة لمربع كاي χ^2 وبدرجات حرية مختلفة.

وممّا تجدر الإشارة إليه هو أنّ إحصائية بيرسون تمثل متغيراً عددياً منفصلاً بينما يمثل χ^2 متغيراً عشوائياً متصلاً وأنّ دالتي الكثافة والتوزيع التكراري لهذه الإحصائية تعتبر من الدوال المتصلة. وتفيد هذه الإحصائية (4.13) كمؤشر للدلالة على مدى التقارب بين نتائج التجربة

الإحصائية الميدانية والقيم المتوقعة أو المفترضة لها فعندما تكون χ^2 صغيرة جداً فإنّ ذلك يعني تقارباً جيداً بين المشاهدات والتوقعات وبالعكس عندما تكون χ^2 كبيرة فإنّ ذلك يعني تباعداً واختلافاً قد يكون جوهرياً وغير مقبول بين المشاهدات الميدانية والتوقعات الحسابية.

4-5-1 خصائص توزيع مربع كاي:

1- $\chi^2 \geq 0$

2- دالة الكثافة $f(\chi^2)$ تعتمد على متغير واحد فقط هو درجة الحرية $v = k - 1$ حيث k عدد المشاهدات أو المعلومات.

3- $E(\chi^2) = v$

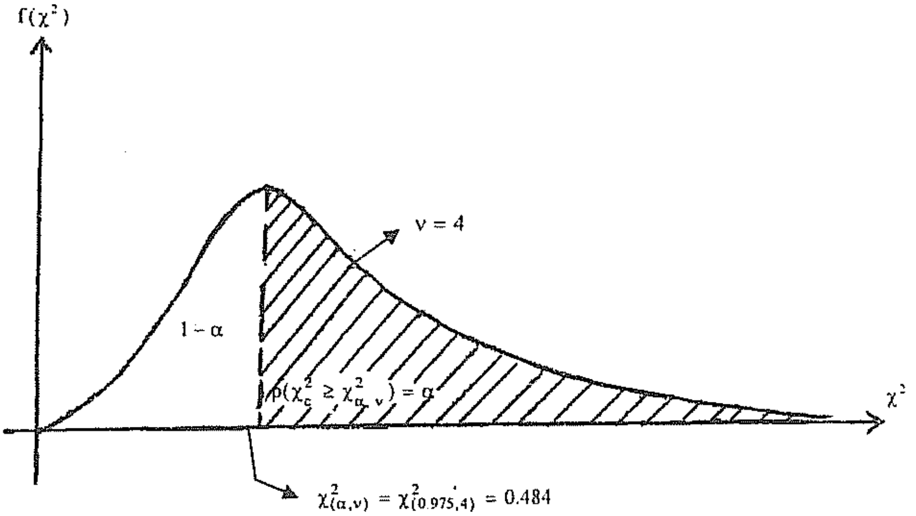
4- $Var(\chi^2) = 2v = 2\mu$

5- القيمة العظمى للدالة تحدث عند النقطة $\chi^2 = v - 1$ وعندما $v > 2$

6- منحني الدالة $f(\chi^2)$ ملتوٍ نحو اليمين ولكنه ينحو نحو التماثل كلما زادت درجات الحرية v

4-5-2 جداول مربع كاي:

يوضح جدول مربع كاي الدرجات الحرجة المناظرة لدرجات الحرية v واحتمالاتها حيث تكتب الاحتمالات في رؤوس الأعمدة بينما تكتب درجات الحرية v بصورة رأسية أمام الصفوف، حيث يمثل الاحتمال المذكور المساحة تحت المنحني وعلى يمين النقطة الحرجة χ^2_v.



الشكل (4-14)

ويتلخص أسلوب استخدام الجداول بحساب احتمال (α) وقوع مربع كاي المحسوب χ_c^2 والأكبر من الدرجة χ_v^2من رأس العمود المتقاطع مع السطر v الذي يمثل درجة الحرية أي حساب $P(\chi_c^2 \geq \chi_v^2)$ والذي يمثل المساحة تحت المنحني في الشكل (4-14). وبالعكس تحسب الدرجة الحرجة $\chi_{\alpha,v}^2$ من تقاطع العمود α مع السطر v.

مثال (13): أحسب كلاً من:

أ – الدرجة الحرجة لمربع كاي $\chi_{(\alpha,v)}^2 = \chi_{(0.05,9)}^2$ وكذلك $\chi_{(0.975,4)}^2$

ب- احتمال $\chi_c^2 > 18.307$

جـ- احتمال $\chi_c^2 \leq 18.307$

الحل: نلاحظ من جدول مربع كاي وفي السطر الموافق للدرجة $v=9$ وتحت العمود $\alpha=0.05$ بأنّ: $\chi_{(0.05,9)}^2 = 16.919$ وبالمثل نجد من الجدول وفي السطر المقابل للدرجة $v=4$ وتحت العمود $\alpha=0.975$ بأنّ $\chi_{(0.975,4)}^2 = 0.484419$.

ب- من الجدول نجد أنّ الرقم 18.307 يقع عند تقاطع العمود $\alpha=0.05$ ودرجة الحرية $v=10$ أي أنّ: $P(\chi^2 > 18.307) = \alpha = 0.05$

جـ- من الفقرة ب نجد أنّ: $P(\chi^2 < 18.307) = 1 - \alpha = 0.95$

4-6 توزيع فيشر (F-distribution):

في الكثير من التجارب الإحصائية نحتاج إلى معرفة إمكانية التغير في مستويات الإنتاج مستقبلاً أو بعد فترات معينة. ففي مضمار الجودة وضبط الإنتاج مثلاً يحتاج مهندس ضبط الجودة لمعرفة ما إذا كانت جودة الإنتاج تتغير بعد فترة زمنية أم لا كما يحتاج الاقتصادي لمعرفة ما إذا كان تباين مردود أو دخول الأفراد من مستوٍ واحد في مجتمعين إحصائيين سيختلف باختلاف مجتمعيهما أم لا أو كما يحتاج الإداري لمعرفة ما إذا كان تباين ساعات العمل في مجتمعي عمال مختلفين يختلف أو يتغير باختلاف المجتمعين أم لا...إلخ.

وللإجابة عن هذه التساؤلات قام فيشر (R.A Fisher ، 1924) بدراسة الموضوع وذلك عن طريق أخذ عينات عشوائية عديدة من كلٍ من مجتمعين إحصائيين مختلفين ولكنهما طبيعيي التوزيع. وكانت الدراسة مبنية على أساس مقارنة تناسب تبايني عينتين مختلفتين من المجتمع

الأول والثاني S_{1i}^2 ، S_{2i}^2 على الترتيب مع تناسب تبايني المجتمعين الأصليين σ_1^2 ، σ_2^2 على الترتيب أيضاً حيث i يمثل رقم العينة المأخوذة من المجتمع الأول والثاني أي مقارنة التناسب:

$$\frac{S_{1i}^2}{S_{2i}^2} \sim \frac{\sigma_1^2}{\sigma_2^2} \sim 1 \qquad \textbf{(4.14)}$$

حيث أنّ التناسب $\frac{S_{1i}^2}{S_{2i}^2}$ يعتبر تقديراً جيداً وغير منحاز للتناسب $\frac{\sigma_1^2}{\sigma_2^2}$. وبتكرار التجربة وحساب التناسبات مرّات عديدة وجد أنّ تكرارات قيم التناسب $\frac{S_{1i}^2}{S_{2i}^2}$ تمثل توزيعاً تكرارياً غير متماثل حول وسطه μ ولكنه منحرف أو ملتوٍ نحو اليمين (skewed to the right) فكلما اقتربت النسبة $\frac{S_{1i}^2}{S_{2i}^2}$ من الواحد اقترب تبايني المجتمعين الإحصائيين الأم من بعضهما البعض (حالة التساوي $\sigma_1^2 = \sigma_2^2$) وبالعكس. وبعبارة أخرى أن الثقة بتكافؤ المجتمعين أو تساوي تغايرهما وعدم اختلافهما تكون ضئيلة كلما ابتعدت النسبة عن الواحد وعلى هذا الأساس تعرف درجة فيشر الفائية المعيارية بالنسبة التالية مأخوذة بالترتيب:

$$F(\nu_1, \nu_2) = S_1^2\sigma_2^2/S_2^2\sigma_1^2 \qquad \textbf{(4.15)}$$

حيث: S_1^2 تباين العينة العشوائية من المجتمع الأول.

S_2^2 تباين العينة العشوائية من المجتمع الثاني.

$\nu_1 = n_1 - 1$ درجة الحرية للعينة الأولى.

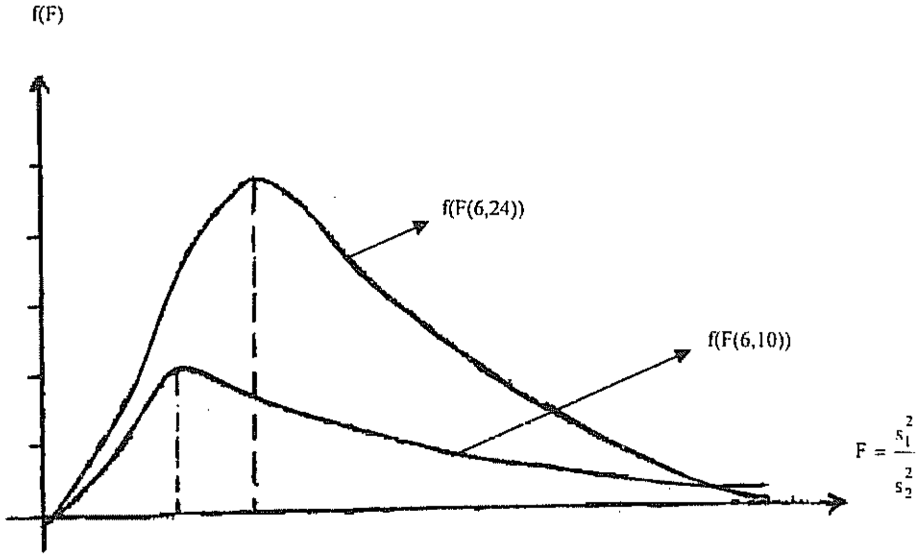
$\nu_2 = n_2 - 1$ درجة الحرية للعينة الثانية.

والدرجة المعيارية الفائية السابقة تمثل متغيراً عشوائياً متصلاً وذو توزيع تكراري يسمى بتوزيع F والذي تعرف دالة كثافته $f(F(\nu_1, \nu_2))$ بالعلاقة التالية:

$$f(F) = \frac{\Gamma\left((\nu_1+\nu_2)/2\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} F^{\left(\frac{\nu_1-2}{2}\right)} \cdot \left(1+\frac{\nu_1}{\nu_2}F\right)^{-\left(\frac{\nu_1-\nu_2}{2}\right)} \qquad \textbf{(4.16)}$$

حيث $0 < F < \infty$ ، والرمز Γ يشير إلى دالة غاما.

كما تأخذ هذه الدالة المنحني التكراري التالي:



الشكل **(4-15)**: توضيح لمنحني دالة الكثافة $f(F(v_1,v_2)$ في الحالتين $(v_1 = 6 \;,\; v_2 = 24)$ و $(v_1 = 6 \;,\; v_2 = 10)$

ومن العلاقة السابقة نلاحظ العلاقة بين كل من دالة التوزيع الاحتمالية لفيشر ودالتي غاما (Γ) وبيتا (β) حيث نعلم أنّ:

$$\Gamma(n) = (n-1)! \,،\, \beta\left(v_1/_2, v_2/_2\right) = \Gamma\left(\frac{v_1 + v_2}{2}\right) \Big/ \left(\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right)\right)$$

ومن خصائص التوزيع السابق F نجد أنّ القيمة المتوقعة للتوزيع:

$$E(F(v_1, v_2)) = \frac{v_2}{v_2 - 2} \qquad \textbf{(4.17)}$$

وأنّ تباينه هو:

$$Var(F(v_1, v_2)) = \frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)} \qquad \textbf{(4.18)}$$

ومنهما نجد أنّ:

$$v_2 \to \infty \quad \text{عندما} \quad Var\,F \to \frac{2}{v_1} \quad , \quad E(F) \to 1$$

الشكل (4-15) يوضح منحني توزيع F بدرجتي حرية $(v_1, v_2) = (6,24)$ ومنحني توزيع F بدرجتي حرية $(v_1, v_2) = (6,10)$ كما يوضح درجة التواء كلا المنحنيين نحو اليمين فكلما

قلّ حجم العينة الثانية زاد التواء المنحني نحو اليمين وبالعكس. والجدير بالملاحظة هو أنّ توزيع F لا يعتمد فقط على درجتي الحرية (v_1, v_2) بل يعتمد على ترتيبهما أيضاً حيث نجد أنّ: $F_{v_1 v_2} \neq F_{v_2 v_1}$

4-6-1 استخدام جداول F:

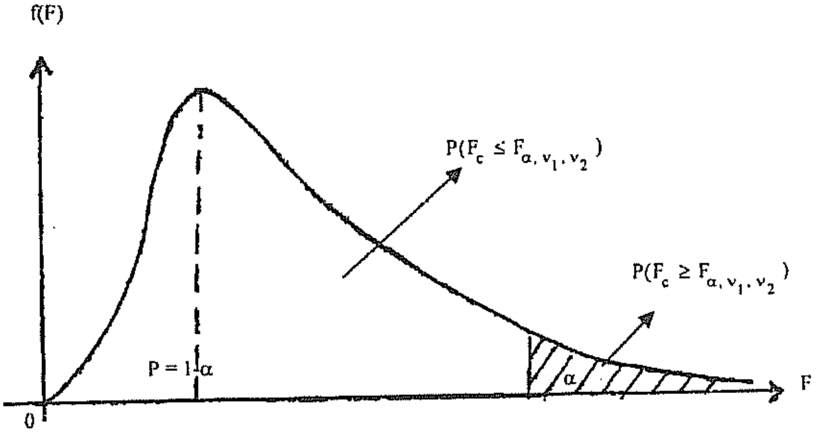
ونظراً لأهمية توزيع فيشر في التحليل الإحصائي بين مجتمعين مختلفين وطبيعيي التوزيع تم جدولة قيم F وحسبت قيمة دالة التوزيع الاحتمالية (دالتي الكثافة والاحتمال) وفقاً لدرجتي الحرية (v_1, v_2) حسب الترتيب. حيث يتم حساب الدرجة المعيارية الفائية F_{α, v_1, v_2} بعد تحديد قيمة α التي تمثل الاحتمال (أو المساحة تحت المنحني) بعد الدرجة الفائية أي على يمين الدرجة الفائية عند تقاطع عمودي درجتي الحرية v_2, v_1 حيث نجد أنّ:

$$F_{\alpha, v_1, v_2} = F_{0.05, 6, 10} = 3.22$$

أي أنّ: $P(F_c \geq 3.22) = 0.05$

وكذلك نجد: $F_{\alpha, v_1, v_2} = F_{0.01, 6, 10} = 5.39$

أي أنّ: $P(F_c \geq 5.39) = 0.01$



الشكل **(4-16)**: يوضح الدرجة الحرجة المعيارية لفيشر F_{α, v_1, v_2} حيث تمثل α احتمال أن تكون النسبة $\frac{S_1^2}{S_2^2}$ أكبر من الدرجة الحرجة

أيّ: $P\left(\frac{S_1^2}{S_2^2} \geq F_{\alpha,(v_1,v_2)}\right) = \alpha$ وانّ $P\left(\frac{S_1^2}{S_2^2} \leq F_{\alpha,(v_1,v_2)}\right) = 1-\alpha$

هنالك جدولان شائعان فقط لتوزيع F ويحسبان قيم الدرجات الفائية المختلفة ضمن احتمالي دالتي التوزيع التجميعية $P(F_c \geq F_\alpha) = 0.5$ لمختلف درجات الحرية v_1, v_2 وكذلك $P(F_c \geq F_\alpha) = 0.01$ لنفس درجات الحرية السابقة وبنفس الترتيب حيث F_c ترمز للدرجة الفائية المحسوبة من التجربة الإحصائية بينما ترمز F_α للدرجة الفائية الحرجة الجدولية. وفي هذين الجدولين جرت العادة على حساب النسب للقيم التي تزيد عن الواحد فقط أي عندما $S_1^2 > S_2^2$ واستخدام الجناح الأيمن للتوزيع أو المنطقة الحرجة اليمنى والتي يمكن أن تحسب احتمال $P\left(\frac{S_1^2}{S_2^2} > F_\alpha\right)$ من الجداول مباشرة أمّا إذا كانت $S_1^2 < S_2^2$ أي أنّ $F_c < 1$ فيتوجب استخدام الجناح الأيسر أو الأدنى للتوزيع والتي تسمى بالمنطقة الحرجة اليسرى وبديهي بمكننا أن نقلب التناسب أي نحسب S_2^2/S_1^2 ومن ثم نحسب $F(v_2, v_1)$ بدلاً من $F(v_1, v_2)$ كما في الحالة الأولى لأنّ النسبة ستكون أكبر من الواحد في هذه الحالة. كما يمكن استخدام العلاقة الخاصة بالقيمتين الحرجتين (critical values) العليا والدنيا والمتعاكستين $F_{\alpha(v_1,v_2)}$ ، $F_{1-\alpha(v_2,v_1)}$ والتي تكتب بالصورة التالية:

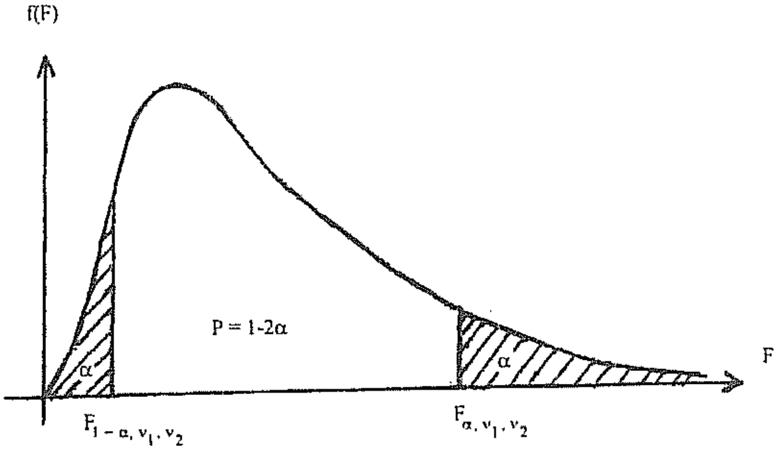
$$F_{1-\alpha, v_1, v_2} = \frac{1}{F_{\alpha, v_2, v_1}} = F^{-1}_{\alpha, v_2, v_1} \qquad \textbf{(4.19)}$$

وباستخدام العلاقة السابقة نجد أنّ:

$$F_{0.95,6,10} = \frac{1}{F_{0.05,10,6}} = \frac{1}{4.06} = 0.246$$

$$F_{0.99,6,10} = \frac{1}{F_{0.01,10,6}} = \frac{1}{7.87} = 0.127$$

الشكل (17-4) التالي يوضح كلاً من الدرجتين الحرجتين العليا والدنيا لفيشر



الشكل (17-4)

يوضح جناحي أو طرفي توزيع F حيث: $P(F_{v_1,v_2} < F_c) = P(F_{v_2,v_1} > \frac{1}{F_c})$

ولحساب احتمال وقوع F ضمن فترة فائية معينة يمكننا استخدام الجداول أو القيم التي تمثل المنحني ذو الطرفين الأيمن والأيسر حيث نجد أنّ:

$$P(F_{1-\alpha,v_1,v_2} < F_c < F_{\alpha,v_1,v_2}) = 1-2\alpha \qquad \textbf{(4.20)}$$

حيث تعتبر F_{α,v_1,v_2} الدرجة الحرجة العليا لفيشر بينما تعتبر $F_{1-\alpha,v_1,v_2}$ الدرجة الحرجة الدنيا لفيشر كما في الشكل السابق (17-4).

والجدير بالذكر من الناحية الرياضية أنّ كلاً من توزيعي F، مربع كاي (χ^2) هما حالات خاصة من بعضهما وكلاهما حالات خاصة من توزيع آخر يسمى بتوزيع غاما (Γ-distribution) وعليه يمكن تعريف الدرجة المعيارية لفيشر بالعلاقة التالية المأخوذة عن التناسب (14-4):

$$F_{v_1,v_2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\chi_1^2/v_1}{\chi_2^2/v_2} \qquad \textbf{(4.21)}$$

أي أن الدرجة المعيارية لفيشر هي عبارة عن النسبة بين مربعي كاي لكلا المجتمعين الإحصائيين مقسومين كلّ على درجة حرية عينته العشوائية كما تنص عليه النظرية التالية:

نظرية ❸:

إذا كانت كل من S_1^2 ، S_2^2 تبايني عينتين عشوائيتين مستقلتين ذواتي حجمين n_1 ، n_2 مأخوذتين من مجتمعين إحصائيين طبيعيي التوزيع وذوي تباينين σ_1^2 ، σ_2^2 على الترتيب فإنّ النسبة:

$$F = \frac{\chi_1^2/v_1}{\chi_2^2/v_2} \qquad (4.22)$$

عبارة عن متغير عشوائي ذي توزيع F وبدرجتي حرية $v_1 = n_1 - 1$ ، $v_2 = n_2 - 1$

مثال (14): في تجربة لدراسة مدخول الأفراد ضمن مجتمعين إحصائيين طبيعيي التوزيع حيث $\sigma_1^2 = \sigma_2^2$، تم أخذ عينتين ذواتي حجمين $n_1 = 11$ ، $n_2 = 16$ وكان تباين كل من العينتين $S_1^2 = (2000)^2$ $S_2^2 = (2400)^2$ على الترتيب. أحسب كلاً من:

$F_{0.95,10,15}$ $F_{0.99,10,15}$ $F_{0.05,10,15}$ $F_{0.01,10,15}$

$F_{0.95,15,10}$ $F_{0.99,15,10}$ $F_{0.05,15,10}$ $F_{0.01,15,10}$

الحل: من الجداول الخاصة نجد أنّ الدرجة الحرجة لفيشر في كلا الجدولين:

$F_{0.05,10,15} = 2.54$ وأن: $F_{0.01,10,15} = 3.80$

وفي كلا الحالتين نجد أنّ الدرجة الفائية المحسوبة من التجربة

$$F_c = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} = \frac{S_1^2}{S_2^2} = \frac{(2000)^2}{(2400)^2} = 0.69$$

هي أصغر من الدرجتين الفائيتين الحرجتين السابقتين وعليه نجد أنّ احتمالية F_c عالية جداً أي أنّ التناسب أو التساوي بين تبايني المجتمعين موجود ومقبول أي أنّ:

$$P(F_c \le F = 3.80) = 1 - \alpha = 0.99$$

$$P(F_c \le F = 2.54) = 1 - \alpha = 0.95$$

ومن العلاقة (4.19) العكسية نجد أنّ:

$$F_{0.99,10,15} = \frac{1}{F_{0.01,15,10}} = \frac{1}{4.56} = 0.22$$

$$F_{0.95,10,15} = \frac{1}{F_{0.05,15,10}} = \frac{1}{2.85} = 0.35$$

$$F_{0.99,15,10} = \frac{1}{3.80}$$
$$F_{0.95,15,10} = \frac{1}{2.54}$$

وبالمقارنة نجد من الحالة الأولى:

$$P(F_{0.99,10,15} = 0.22 < F_c = 0.69 < 3.80 = F_{0.01,10,15}) \geq 0.98 = 1 - 2\alpha$$

ومن الحالة الثانية:

$$P(F_{0.95,10,15} = 0.35 < F_c = 0.69 < 2.54 = F_{0.05,10,15}) \geq 0.90 = 1 - 2\alpha$$

4-7 تمارين عامة:

1- إذا علمت أنّ توزيع X هو توزيع طبيعي متوسطه $\mu=12$ وانحرافه المعيـــاري $\sigma = 2$ فأحسب مستخدماً الجداول كلاً من الاحتمالات التالية:

$P(X > 14)$	$P(X < 10)$	$P(10 < X < 13)$
$P(X > 11)$	$P(X < 10.5)$	$P(16 < X < 20)$

2- إذا كانت درجتك الطبيعية المعيارية في إحدى الاختبارات هي $z = 8$ فما هـــي النســـبة المئوية للطلبة الذين تتوقع أن تكون درجاتهم المعيارية أعلى من درجتك.

3- أحسب كلاً من الوسط والانحراف المعياري لتوزيع متغيـــر ذي الحـــدين إذا علمـــت أنّ $n = 12 \qquad P = \frac{1}{4}$

4- إذا كان متوسط الزمن بين دخول سيارة وأخرى لمحطة بنزين هو 5.5 دقيقة فحدد نوعية التوزيع وأحسب كلاً من متوسطه μ وتباينه σ^2 .

5- أُخذت عينتان على الترتيب من مجتمعين طبيعيي التوزيع فما هو نوع توزيع النسبة بـــين تباينهما، وإذا كان حجم العينة الأولى $n_1 = 31$ وحجم العينة الثانية $n_2 = 25$ فأحسب كلاً من: $P(A \geq 2.47)$, $P(A < 2.47)$

6- عائلة عندها 10 أطفال. بفرض أنّ للصبي والبنت نفس الاحتمال في الوجود.

أ- ما هو احتمال وجود 5 صبيان و5 بنات.

ب- ما هو احتمال وجود من 3 وحتى 7 صبيان.

7- أنتج مصنع للأدوية دواءً جديداً لمعالجة أحد الأمراض، معدل نجاحه 80%، أُعطي هـــذا الدواء لـــ 15 مريضاً بهذا المرض. ما احتمال شفاء 12 منهم على الأقل؟

8- آلة لإنتاج المسامير لوحظ أنها تنتج مسامير تالفة بنسبة 20% اخترنا عشوائياً 10 مسامير من المسامير المنتجة من هذه الآلة والمطلوب:

أ- ما احتمال وجود مسمارين تالفين فقط بين هذه المسامير.

ب- ما احتمال وجود ثلاثة مسامير تالفة على الأكثر.

9- إذا كان معدّل حدوث أعطال كهربائية في منطقة ما في اليوم هو 6 والمطلوب:
أ- ما احتمال عدم حدوث أي عطل كهربائي في يوم معين.
ب- ما احتمال حدوث عطلين كهربائيين على الأكثر في يوم معين.

10- ينتج معمل للبطاريات، بطاريات صغيرة متوسط أعمارها 30 ساعة بانحراف معياري قدره 5 ساعات، وحيث أنّ أعمار البطاريات تتوزع وفق التوزيع الطبيعي، فما هو احتمال أن تعيش بطارية ما أقل من 23 ساعة؟

11- إنّ درجات 300 طالب نجحوا في امتحان مقرر الإحصاء كانت تتوزع طبيعياً، بمتوسط قدره 65 وبانحراف معياري قدره 10 والمطلوب:
أ- كم هو عدد الطلاب الذين وقعت درجاتهم بين السبعين والثمانين.
ب- ما هي أقل درجة حصل عليها طالب من الـ 12% الأوائل، وما هي أعلى درجة حصل عليها طالب من المجموعة الباقية.

12- رميت قطعة نقود متوازنة سبع مرّات فما هي نوعية توزيع الحصول على الصورة H. وما هو احتمال الحصول على ثلاث صور، أربع صور، لا صورة، وهل يكون التوزيع في هذه الحالة طبيعياً.

13- أكتب ووضح العلاقة بين توزيع فيشر وتوزيع مربع كاي.

14- فصل دراسي فيه 30 طالباً فإذا كانت نسبة النجاح 60% فحدّد نوعية التوزيع وأحسب كلاً من متوسط وتباين التوزيع ثم أكتب عنصراً واحداً من عناصر فضاء التجربة.

15- إذا كان S_1^2،S_2^2 تبايني عينتين عشوائيتين ذواتي حجم $n_1 = 25$ ، $n_2 = 31$ مأخوذتين من مجتمعين إحصائيين تباينهما $\sigma_1^2 = 10$ ، $\sigma_2^2 = 15$ على الترتيب، فما هو نوع توزيع النسبة بين التباينين وما هو احتمال $S_1^2/S_2^2 > 1.26$ ، $S_1^2/S_2^2 < 1.26$

16- إذا كان حجم العينتين $n_1 = 8$ ، $n_2 = 12$ فما هو احتمال $S_1^2/S_2^2 < 4.89$ على افتراض تساوي تبايني المجتمعين الإحصائيين.

الفصل الخامس

نظرية التقدير

5-1 – شروط التقدير.

5-2 – التقدير النقطي والمجالي للوسطاء.

5-3 – تقدير متوسط المجتمع الإحصائي μ.

5-4 – تقدير تباين المجتمع الإحصائي σ^2.

5-5 – تقدير النسبة بين تبايني مجتمعين إحصائيين $\frac{\sigma_1^2}{\sigma_2^2}$.

5-6 – التمثيل الهندسي (البياني) لمجال الثقة.

5-7 – تقدير الاحتمال P للمجتمع الإحصائي.

5-8 – تقدير معلمة الفرق بين احتمالين P_D لمجتمعين إحصائيين.

5-9 – تمارين عامة.

نظرية التقدير
(Estimation theory)

تعتبر نظرية التقدير لوسطاء المجتمع الإحصائي أحد أهم فروع التحليل الإحصائي بعد تحليل التباين واختبار الفرضيات. وتتميز عمليات التقدير بالكثير من الأخطاء الرقمية والفكرية وأخطاء التقريب والتدوير علاوة على تعددية التقديرات. فمثلاً نلاحظ أنّ كلاً من متوسط العينة $\overline{X}$ والمنوال M يمكن أن يعتبرا دوالاً لتقدير متوسط المجتمع الإحصائي μ خصوصاً إذا كان توزيع المتغير X طبيعياً وبالمثل نلاحظ أنّ كلاً من العلاقتين التاليتين لتباين العينات:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \qquad\qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

يمكن اعتبارهما إحصائيات لتقدير تباين المجتمع الإحصائي σ^2.

وعلى سبيل المثال يتم تقدير العمر الزمني للبرادات الكهربائية كافة بمتوسط هذا العمر والذي يرمز له بالرمز μ حيث يتم حساب هذا التقدير على أساس متوسط العمر الزمني $\overline{X}$ لعينة عشوائية مناسبة وغير منحازة من تلك البرادات. وهكذا بالنسبة لكل من التباين والانحراف المعياري للمجتمع.

ومن الجدير بالذكر أنّ نظرية التقديرات الوسيطية تعتمد على نوعية التوزيع التكراري لهذا الوسيط وهو ما يؤكد ضرورة تحديد ما إذا كان التوزيع معتدلاً (طبيعياً) أم لا. كما نود الإشارة إلى أنّ هنالك العديد من طرق تقدير الوسطاء نذكر منها على سبيل المثال: طريقة المربعات الصغرى (least square) وطريقة الترجيح الأعلى (Maximum likelihood) وطريقة العزوم (Moments) وطريقة بايز (Bayes) وغيرها... إلاّ أننا سنقتصر على أسلوب التوزيعات التكرارية لتقدير هذه الوسطاء بعد معرفة شروط ومستلزمات التقدير:

5-1- شروط التقدير:

تتميز نظرية التقديرات بشروطها المتعددة والتي نذكر منها التالي:

1- الدقة وعدم الانحياز (accurecy and un baisedness):

يُعرّف عدم الانحياز بتساوي الوسيط المراد تقديره θ مثلاً مع توقع وسيط التقدير $\hat{\theta}$ أيّ أنّ مقدار الخطأ في التقدير يساوي صفراً وبعبارة رياضية نجد أنّ شرط عدم الانحياز هو:

$$E(\hat{\theta}) = \theta \qquad (5.1)$$

ويعتبر شرط عدم التحيز من أهم شروط التقديرات الوسيطية.

مثال (1): وضّح بأنّ الإحصائية $a = \frac{n_1 s_1^2 + \ldots + n_k s_k^2}{r}$ حيث k يرمز لعدد العينات بينما $r > 0$ عدد حقيقي، هي وسيط تقدير منحاز لتباين المجتمع σ^2 وأوجد التعديل اللازم لها كي تكون وسيط غير منحاز.

الحل:

$$E(a) = E\left(\frac{n_1 s_1^2 + \ldots + n_k s_k^2}{r}\right) = \frac{1}{r}[n_1 E(s_1^2) + \ldots + n_k E(s_k^2)]$$

$$= \frac{1}{r}[n_1\sigma^2 + \ldots + n_k\sigma^2] = \frac{\sigma^2}{r}\sum_{i=1}^{k} n_i \neq \sigma^2$$

ومن العلاقة الأخيرة نجد أنّ مقدار الانحياز هو:

$$B = E(a) - \sigma^2 = \frac{\sum n_i}{r}\sigma^2 - \sigma^2 = \sigma^2\left(\frac{\sum n_i}{r} - 1\right) \neq 0$$

ولأجل أن تكون a غير منحازة يتوجب أن يكون $B = 0$ أو أن:

$$\left(\sum_{i=1}^{k} n_i \Big/ r\right) - 1 = 0 \qquad \text{أي أنّ:} \qquad r = \sum_{i=1}^{k} n_i$$

وهو الشرط اللازم للتعديل كي تكون a غير منحازة والتي تأخذ الصورة التالية:

$$a = \frac{n_1 S_1^2 + n_2 S_2^2 + \cdots + n_k S_k^2}{n_1 + n_2 + \cdots + n_k}$$

مثال (2): إذا كانت $X_1, \ldots\ldots, X_6$ هي عناصر عشوائية مستقلة من مجتمع إحصائي طبيعي التوزيع بمتوسط μ، وتباين σ^2.

أحسب قيمة c بحيث يكون الوسيط:

$$T = C[(X_1 - X_2)^2 + (X_3 - X_4)^2 + (X_5 - X_6)^2]$$

غير منحاز لتباين المجتمع σ^2.

الحل: ليكون الوسيط T غير منحاز يتوجب أن يكون:

$$E(T) - \sigma^2 = 0$$

$$E(T) = E[C\{(X_1 - X_2)^2 + (X_3 - X_4)^2 + (X_5 - X_6)^2\}]$$

بإضافة μ وطرح μ.

$$E(T) = CE\{[(X_1 - \mu) - (X_2 - \mu)]^2 + [(X_3 - \mu) - (X_4 - \mu)]^2 + [(X_5 - \mu) - (X_6 - \mu)]^2\}$$

$$= CE\{\sum_{i=1}^{6}(X_i - \mu)^2 - 2(X_1 - \mu)(X_2 - \mu) - 2(X_3 - \mu)(X_4 - \mu) - 2(X_5 - \mu)(X_6 - \mu)\}$$

$$= C\{\sum_{i=1}^{6}E(X_i - \mu)^2 - 2[E(X_1 - \mu)(X_2 - \mu) + E(X_3 - \mu)(X_4 - \mu) + E(X_5 - \mu)(X_6 - \mu)]\}$$

$$= C\ \{\sum_{i=1}^{6}E(X_i - \mu)^2 - 2[\mathrm{cov}(X_1, X_2) - \mathrm{cov}(X_3, X_4) - \mathrm{cov}(X_5, X_6)]\} = \sigma^2$$

ومنه نجد أنّ:

$$C = \frac{\sigma^2}{\sum_{1}^{6}\sigma^2 - 2\sum_{i=1,3,5}\mathrm{cov}(X_i X_{i+1})}$$

وبما أنّ: $\sum_{i=1}^{6}\sigma^2 = 6\ \sigma^2$ و $(i \neq j)$ و $\mathrm{cov}(X_i, X_j) = 0$

فإنّ: $C = \dfrac{\sigma^2}{6\sigma^2} = \dfrac{1}{6}$

مثال (3): إذا كانت $X_1, \ldots\ldots, X_n$ هي عناصر عشوائية مستقلة من مجتمع إحصائي ذي توزيع بواسوني وبفرض الوسيط M ، المعرّف بالشكل التالي:

$$M = d[X_1(X_1 - X_2) + X_2(X_2 - X_3) + \ldots\ldots + X_{n-1}(X_{n-1} - X_n)]$$

أحسب قيمة d بحيث يكون M غير منحاز في تقدير التباين للمجتمع الإحصائي:

الحل:

$$E(M) = dE[X_1(X_1 - X_2) + X_2(X_2 - X_2) + \ldots\ldots + X_{n-1}(X_{n-1} - X_n)]$$

$$= dE\ \ [X_1^2 - X_1X_2 + X_2^2 - X_2X_3 + \ldots + X_{n-1}^2 - X_{n-1}X_n]$$

$$= dE[\sum_{i=1}^{n-1}X_i^2 - \sum_{i=1}^{n-1}X_iX_{i+1}]$$

$$= d\sum_{i=1}^{n-1}[E(X_i^2) - E(X_i X_{i+1})]$$

$$= d\sum_{i=1}^{n-1}[E(X_i^2) - E(X_i)E(X_{i+1})]$$

$$= d\ \ (n-1)\ \ [\sigma^2 + \mu^2 - \mu^2] = \sigma^2$$

$$d\ (n-1)\ \sigma^2 = \sigma^2$$

$$d = \frac{\sigma^2}{(n-1)\sigma^2} = \frac{1}{n-1}$$

2- الكفاءة (efficiency):

يُقال عن التقدير $\hat{\theta}_1$ بأنه أكثر كفاءة من التقدير $\hat{\theta}_2$ في تقدير الوسيط θ إذا كانا:

أ- تقديرين غير منحازين أصولاً.

ب- $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ وهذا يعني أنّ نزعة التمركز للمقدّر $\hat{\theta}_1$ حول θ أقوى منه في الثانية. كما يُقال عن التقديرين $\hat{\theta}_1$ و $\hat{\theta}_2$ بأنهما متكافئان إذا تساوى تباينهما:

$$Var(\hat{\theta}_1) = Var(\hat{\theta}_2) \qquad \textbf{(5.2)}$$

وهنالك تعاريف أو شروط أخرى للكفاءة منها ما يسمى بالكفاءة النسبية $Var(\hat{\theta}_1)/Var(\hat{\theta}_2)$ ومنها ما يسمى بالكفاءة المقاربة (Asymptotic efficiency) وهما خارج نطاق منهاجنا.

3- الكفاية (sufficiency):

يُقال للمقدر $\hat{\theta}(x_1,...,x_n)$ بأنه مقدر كافٍ لتقدير θ إذا وفقط إذا كان التوزيع التكراري للعينة $(x_1,...,x_n)$ والمشروط بالشرط $\hat{\theta} = k$ لا يعتمد على الوسيط θ ذاته ولكافة قيم k.

والكفاية تعني عدم الحاجة إلى إضافة عناصر أو معلومات أخرى من المجتمع الإحصائي لأن هذه العناصر الإضافية سوف لن تعطي من المعلومات أكثر مما أعطاه المقدر $\hat{\theta}$ نفسه، ولو افترضنا بأن هنالك مقدّراً آخر مثل $\hat{B}(x_1, x_2,...,x_n)$ لتقدير θ فإننا سنجد أنّ $\hat{B}$ يعتمد أو يرتبط بشكل ما بالمقدر $\hat{\theta}$ نفسه وليس مستقلاً عنه.

4- التناسق والتماسك (consistency):

يُقال عن متتالية التقديرات $\{\hat{\theta}_n\}$ لتقدير الوسيط θ بأنها متناسقة إذا كانت متقاربة الاحتمال من الوسيط θ وبعبارة رياضية إذا كان لكل $\varepsilon > 0$

$$\lim_{n\to\infty} P(\ |\hat{\theta}_n-\theta| > \varepsilon\ ) = 0 \qquad \text{أ-} \qquad (5.3)$$

$$\lim_{n\to\infty} P(\ |\hat{\theta}_n-\theta| \le \varepsilon) = 1 \qquad \text{ب-} \qquad (5.4)$$

ويلاحظ بأنّ $\{\hat{\theta}_n\}$ عبارة عن متتالية من الوسطاء أو الإحصائيات لتقدير الوسيط $\hat{\theta}$. ومن جانب آخر يُقال عن المتتالية $\{\hat{\theta}_n\}$ بأنها متناسقة إذا كان تباينها يؤول إلى الصفر عندما $n \to \infty$ كما تنص عليه النظرية التالية:

نظرية ❶ (الشرط الكافي للتناسق):

إذا كانت $\{\hat{\theta}_n\}$ متتالية تقدير غير منحازة فإنّ $\{\hat{\theta}_n\}$ تكون متناسقة إذا كان:

$$\lim_{n\to\infty} Var(\hat{\theta}_n) = 0 \qquad (5.5)$$

5- التراصف والتركيز (closeness):

يُقال عن المقدّر $\hat{\theta}_1$ لتقدير الوسيط θ بأنه أكثر تراصفاً أو تركيزاً من المقدّر $\hat{\theta}_2$ في تقدير θ إذا كان احتمال أو درجة الثقة بالأول أكبر من الثاني أي أنّ:

$$P(\theta-\lambda < \hat{\theta}_1 \le \theta+\lambda) > P(\theta-\lambda < \hat{\theta}_2 \le \theta+\lambda) \qquad (5.6)$$

ولكافة قيم $\lambda > 0$

ومن جانب آخر يمكن استخدام تعريف بتمان (Bitman) في هذا الصدد حيث:

$$P(|\hat{\theta}_2-\theta| < |\hat{\theta}_1-\theta|) \ge \frac{1}{2} \qquad (5.7)$$

5-2 التقدير النقطي والمجالي للوسطاء point and Interval Estimation:

يعرف التقدير النقطي لأي وسيط مثل θ بالتقدير العددي المفرد الذي يحدد قيمة معينة واحدة فقط مثل $\hat{\theta}$ لذلك الوسيط. كأن يُقال بأن متوسط غلاء المعيشة θ في مدينة دمشق هو $\hat{\theta} = 20000$ ليرة سورية.

ويعرّف التقدير المجالي لأي وسيط مثل θ بمجموعة من العناصر الحقيقية التي تحتوي على θ نفسها ويكتب هذا التقدير عادة على صورة المتراجحة أو المتباينة التالية $L \le \hat{\theta} \le U$ حيث U,L يمثلان الحد الأدنى والأعلى للمجموعة على الترتيب. وتصنف فترات أو مجالات التقدير عادةً إلى نوعين هما:

أ – المجال الثابت (Fixed Interval) مثل $1 < \theta < 3$

ب– المجال المتغير غير الثابت (Flexable Interval) مثل $\theta < 7 < 3\theta$ ويمكن تحوير أحد المجالين إلى المجال الآخر لحساب احتمال احتواء مجال ما معين على وسيط التقدير θ أي حساب الاحتمال $P(X < a < bX)$

مثال (4): أ – حوّل الفترة الثابتة $1 < \theta < 3$ إلى الصورة المتغيرة

ب– حوّل الفترة المتغيرة $\theta < 3.3 < 1.2\theta$ إلى الصورة الثابتة

الحل: أ – نلاحظ أنّ الفترة $1 < \theta < 3$ تعني أنّ $\theta < 3$ وكذلك $1 < \theta$ ومنها نجد أنّ $3 < 3\theta$ ومن دمج المتراجحتين الأولى والأخيرة نجد أنّ $\theta < 3 < 3\theta$

ب– من المتراجحة نجد أنّ $\frac{3.3}{1.2} < \theta$ وكذلك $\theta < 3.3$ وبالدمج نجد أنّ $\frac{3.3}{1.2} < \theta < 3.3$ وهي فترة ثابتة.

مثال (5): إذا كانت X عبارة عن متغير ذو توزيع مربع كاي χ^2_{18} ودرجة حرية $v = 18$ فما هو احتمال أن تحتوي الفترة $(X,3.5X)$ على القيمة (أو النقطة) $\chi^2 = 28.9$ وما هي القيمة المتوقعة $E(X)$ لطول الفترة أعلاه.

الحل: نلاحظ أولاً بأنّ:

$$P(X < \chi^2) = \int_0^{\chi^2} \frac{1}{\Gamma(\frac{r}{2})\, 2^{r/2}} \omega^{r/2 - 1} e^{-\omega/2} d\omega$$

نحوّل الفترة المتغيرة إلى فترة ثابتة كالتالي:

$$X < 28.9 < 3.5X$$

وهذا يعني أنّ:

$$\chi^2_{18} \leftarrow L = \frac{28.9}{3.5} \quad < X < 28.9 = U \rightarrow \chi^2_{18}$$

وبمقارنة كل من الحدين الأدنى L والأعلى U مع جدول مربع كاي χ^2_{18} وبدرجة حرية $v = 18$ نحصل على التالي:

$$P(X = \chi^2 < 28.9) = 0.95$$

بينما نجد أنّ احتمال الحد الأدنى L هو:

$$P(X = \chi^2 < 8.257) = 0.025$$

وبطرح الاحتمالين من بعضهما نجد أنّ:

$$P(X < 28.9 < 3.5X) = P(8.257 < X < 28.9)$$

$$= P(X < 28.9) - P(X < 8.257) = 0.925$$

أمّا القيمة المتوقعة لطول الفترة حيث أنّ: $E(X) = v = 18$ فتحسب كالتالي:

$$E(\text{طول الفترة}) = E(3.5X - X) = E(2.5X) = 2.5E(X) = 2.5(18) = 45$$

مثال (6): إذا كانت x_1 ، x_2 عبارة عن مفردات عينة عشوائية مأخوذة من مجتمع ذي توزيع تكراري معيّن وبدالة كثافة هي:

$$f(x) = \begin{cases} \frac{1}{4} & x = 1, 2, 3, 4 \\ 0 & \text{لغير ذلك} \end{cases}$$

فاحسب احتمال احتواء فترة التقدير المتغيرة $(1, x_1 + x_2 - \frac{1}{2})$ على القيمة $x = 3$

الحل: نفرض احتواء الفترة على $x = 3$ أي أنّ:

$$1 < 3 < x_1 + x_2 - \frac{1}{2}$$

ومنها نجد: $1.5 < 3.5 < x_1 + x_2$ أي أنّ $3.5 < x_1 + x_2$

وبأخذ الاحتمال نجد:

$$P(3.5 < X_1 + X_2) = 1 - P(X_1 + X_2 \le 3.5)$$

$$= 1 - \sum_{i=1}^{n} (t_i)\, P(X = t_i) \qquad \text{حيث افترضنا } t_i = x_1 + x_2$$

$$= 1 - \sum_{i=1}^{n} (t_i)\, .\frac{f(t_i)}{n}$$

$$= 1 - \sum (t_i)\, .\frac{1/4}{4} = 1 - \frac{1}{16} \sum t_i = 1 - \frac{1}{16}(1 + 2) = \frac{13}{16}$$

(لاحظ بأنّ $t_i = x_1 + x_2 \le 3.5$ لن تحدث إلاّ عندما $x_1 = 1$ ، $x_2 = 2$ وبالعكس فقط)

مثال (7): احسب احتمال احتواء فترة التقدير المتغيرة $(|x|, 12\ |x|)$ على الانحراف المعياري σ وأحسب القيمة المتوقعة للفترة إذا علمت أنّ X متغير ذو توزيع طبيعي $N(0, \sigma^2)$ ، $\sigma^2 > 0$

الحل: نفترض وقوع الانحراف المعياري σ في فترة التقدير، أي أنّ: $|x| < \sigma < 12\ |x|$
ومنها نجد أنّ: $\sigma < 12\ |x|$ وكذلك $\sigma > |x| > \frac{\sigma}{12}$ ومن المتراجحة الأخيرة نجد أنّ:
$\frac{\sigma}{12} < x < \sigma$ وكذلك $-\sigma < x < \frac{-\sigma}{12}$ وبأخذ الاحتمالات نجد أنّ احتمال كلا الفترتين الموجبة والسالبة متساوي لتساوي طوليهما تحت المنحني الطبيعي. وعليه فإنّ:

$$P\left(|X| < \sigma < 12\,|X|\right) = \frac{-\sigma}{\sqrt{2\pi}} \int_{\sigma/12}^{\sigma} \frac{-x}{\sigma^2} e^{-x^2/2\sigma^2} dx$$

$$= \frac{-\sigma}{\sqrt{2\pi}} \left[e^{-x^2/2\sigma^2}\right]_{\sigma/12}^{\sigma} = \frac{-\sigma}{\sqrt{2\pi}} \left[e^{-1/2} - e^{-1/288}\right]$$

أمّا القيمة المتوقعة فتحسب من التوزيع الطبيعي كالتالي:

$$E(|X|) = 2\int_0^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx = \frac{-2\sigma}{\sqrt{2\pi}} \int_0^{\infty} \frac{-x}{\sigma^2} e^{-x^2/2\sigma^2} dx = \sigma \frac{\sqrt{2}}{\sqrt{\pi}}$$

5-3- تقدير متوسط المجتمع الإحصائي μ :

تعتبر أفضل وسيلة للتقدير النقطي لمتوسط مجتمع إحصائي μ هي استخدام متوسط العينة $\overline{X}$ ويتميز هذا التقدير بعدة خصائص مهمة منها أنه تقدير غير منحاز وذلك لانعدام الخطأ فيه أي: $\varepsilon = E(\overline{X}) - \mu = 0$

ولتوضيح العلاقة السابقة نلاحظ أنّ:

$$E(\overline{X}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n}(E(x_i)) = \frac{1}{n}(n\mu) = \mu$$

ولتقدير المتوسط μ نميز الحالات الثلاث التالية:

1- عندما يكون تباين المجتمع الإحصائي الأم معروفاً.

2- عندما يكون تباين المجتمع الإحصائي الأم غير معروف.

3- عندما يكون المجتمع الإحصائي محدوداً أو منتهياً.

وهي نفس الحالات التي تؤخذ بعين الاعتبار عند تقدير المجموع أو الفرق لمتوسطي مجتمعين إحصائيين $(\mu_1 \pm \mu_2)$.

5-3-1 تقدير المتوسط الحسابي النقطي μ عندما يكون تباين المجتمع σ^2 معروفاً:

يسهل عند استخدام المتوسط الحسابي $\bar{x}$ لعينة عشوائية مأخوذة من مجتمع إحصائي ذي تباين σ^2 معروف، لتقدير متوسطة μ، حساب مقدار الخطأ النسبي z في التقدير (Error of Estimation) وذلك باعتماد نظرية الحد المركزي (Central limit theorem) حيث نجد منها أنّ:.

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \qquad (5.8)$$

حيث يمثل البسط مقدار الخطأ الأكبر ε بينما يمثل Z متغيراً عشوائياً ذا توزيع طبيعي معياري. ومن العلاقة أعلاه يمكننا تقدير قيمة μ وبدرجة ثقة $p = 1-\alpha$ حيث α تمثل مقدار الاحتمال المتطرف من أقصى جهتي اليمين واليسار للتوزيع الطبيعي المعياري:

$$\varepsilon = \bar{x} - \mu = z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right) \qquad (5.9)$$

حيث $z_{\alpha/2}$ هي قيمة المتغير الطبيعي المعياري (الدرجة المعيارية الحرجة) التي تقع على يمينها مساحة تساوي $\alpha/2$. ومن الجدير بالانتباه في الحسابات العادية هو أنّ الاحتمال $P = 1-\alpha$ الشائع في هذا الصدد هو $P = 0.95$ ، $P = 0.99$ أي أنّ $\alpha = 0.05$ ، $\alpha = 0.01$ على الترتيب.

ومن الجداول الخاصة بالتوزيع التكراري الطبيعي المعياري نجد أنّ: $z_{\alpha/2} = z_{0.025} = 1.96$ ، $z_{\alpha/2} = z_{0.005} = 2.575$ على الترتيب أيضاً. وما تعنيه هذه النسب هو أننا على درجة $P = 0.95$ ، أو $P = 0.99$ من الثقة بأنّ أعلى خطأ في تقدير $\bar{X}$ لمتوسط المجتمع الإحصائي μ لا يتجاوز أبداً المقدار $\varepsilon_{\bar{x}} = z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$ وإذا ما استطعنا

الحصول على مقدّر آخر مثل $\hat{\theta}$ بحيث أنّ $\varepsilon_{\hat{\theta}} < \varepsilon_{\bar{x}}$ وبنفس درجة الثقة $P = 1 - \alpha$ فإنّ من الأفضل اختيار $\hat{\theta}$ بدلاً من $\overline{X}$ لتقدير متوسط المجتمع μ.

ومن الجدير بالذكر أيضاً هو أنّ العلاقة (5.8) قد تستخدم في عدة مجالات علاوة على تحديد مقدار الخطأ نذكر منها تحديد حجم الثقة أي حساب α أو تحديد حجم العينة العشوائية n أو حساب وتقدير الانحراف المعياري للمجتمع σ.

مثال (8): يرغب أحد خبراء الكفاءة باستخدام المتوسط الحسابي لعينة عشوائية ذات حجم $n = 150$ لتقدير متوسط الكفاءة الميكانيكية لعدد معين من العمال في إحدى الصناعات الكبيرة للسيارات. فإذا افترض وبناءً على الخبرة بأن $\sigma = 6.2$ فكم يكون أعلى خطأ في تقدير المتوسط الحسابي μ للمجتمع العمالي على أساس 0.99 من الثقة. وإذا افترضنا بأنّ المتوسط الحسابي للعينة السابقة $\bar{x} = 69.5$ فكم هو تقدير المتوسط الحسابي μ لكفاءة العمال كافة في تلك الصناعة وذلك بنفس الدرجة 0.99 من الثقة.

الحل: لدينا $1 - \alpha = 0.99$ ومنه $\alpha = 0.01$ أي $\alpha/2 = 0.005$ عندئذٍ بالتعويض في قانون الخطأ الأكبر ε المبين في العلاقة (5.9) نجد أنّ:

$$\varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = z_{0.005} \frac{6.2}{\sqrt{150}} = 2.575 \frac{6.2}{\sqrt{150}} = 1.30$$

وتعني هذه النتيجة بأن الخبير يستطيع الجزم وبدرجة 0.99 من الثقة بأنّ أعلى مقدار للخطأ في التقدير هو 1.30 وعلى هذا الأساس نجد أنّ تقدير المتوسط الحسابي للمجتمع (الكفاءة العمالية) هو: $\mu = \bar{x} \pm 1.30$ ولو افترضنا بأنّ $\bar{x} = 69.5$ من الدراسة الميدانية فإنّ التقدير (النقطي) لمتوسط المجتمع هو: $\mu_1 = 68.20$ أو $\mu_2 = 70.80$.

مثال (9): كم يجب أن يكون حجم العينة العشوائية n التي يجب أن تؤخذ من مجتمع إحصائي للمحركات لتحديد المتوسط الزمني الذي تستغرقه آلة معينة في تحريك عجلة سيارة على أساس 95٪ من الثقة بأنّ أعلى خطأ في تقدير المعلمة μ هو 0.50 دقيقة وإذا علمت، من تجارب سابقة ميدانية، بأنّ تباين المجتمع الإحصائي معروف ويساوي $\sigma^2 = 2.56$ دقيقة.

الحل: بتطبيق قانون الخطأ الأكبر في التقدير (5.9) نجد أنّ:

$$n = \left(\frac{1.96}{0.5}\right)^2 .(2.56) = 39.3 \cong 40$$

أي أننا يجب أن نفحص 40 محركاً من مجموع كافة المحركات على الأقل كي يكـون أكبـر خطأ في التقدير هو 0.50

ملاحظة ①: يمكن عكس المسألة وتعديلها للحصول على التقدير النقطي μ لكافة المحركـات بفرض أنّ: $n = 40$ حيث نجد أنّ:

$$\mu = \overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \overline{x} \pm z_{0.025} \frac{106}{\sqrt{40}} = \overline{x} \pm 1.96 \frac{106}{\sqrt{40}}$$

وعلى افتراض أنّ متوسط زمن التحريك للعينة $\overline{x} = 2{,}30$ دقيقـة نجـد أنّ تقـدير متوسـط المجتمع النقطي هو: $\mu = 2.30 \pm 0.496$ أي أنّ: $\mu_1 = 2.796$ و $\mu_2 = 1.804$

5-3-2 تقدير المتوسط الحسابي النقطي μ للمجتمع الإحصائي عندما يكون σ^2 غير معروف:

في هذه الحالة يمكننا الاستعاضة عن تباين المجتمع الإحصائي σ^2 غير المعروف بتباين العينة العشوائية S^2 المعروف وذلك شريطة أن يكون التوزيع التكراري للمجتمع الإحصـائي مـن النوع الطبيعي وإن لم يكن كذلك فهنالك أساليب أخرى مختلفة يمكن الرجوع إليها في مصادر أخرى.

وعلى افتراض أنّ التوزيع من النوع الطبيعي فإنّ المتغير العشـوائي المسـتخدم الجديـد (أو درجة الانحراف النسبي) سيكون متغير ستيودنت t المعياري. ويحسب من العلاقة التالية:

$$t = \frac{\overline{x} - \mu}{S/\sqrt{n}} \qquad \textbf{(5.10)}$$

حيث t هي قيمة المتغير في توزيع ستيودنت ذو $v = n - 1$ من درجات الحرية، وباسـتخدام الاختبار (t) أعلاه يمكننا الجزم وبنسبة $p = 1 - \alpha$ من درجات الثقة بأنّ أعلى مقدار للخطـأ في هذا التقدير يساوي:

$$\varepsilon = \overline{x} - \mu = t_{\alpha/2} . \frac{S}{\sqrt{n}} \qquad \textbf{(5.11)}$$

ومن العلاقة أعلاه يمكن تقدير متوسط المجتمع الإحصائي μ بالاعتماد على $\overline{x}$ وبدرجة مـن الثقة $p = (1 - \alpha)\ \%$.

مثال (10): من خلال ست عمليات تجريبية لتحديد درجة انصهار مادة معينة، لوحظ بأنّ المتوسط الحسابي لدرجة الانصهار هي $\bar{x} = 232.26$ درجة وبانحراف معياري مقداره $S = 0.14$ درجة فما هو مقدار أكبر أو أعلى خطأ في تقدير المتوسط الحسابي للمجتمع إذا علمت أن درجة الثقة المطلوبة هي: 98٪.

الحل: من المعلومات أعلاه نجد أنّ: $\alpha = 0.02\,, S = 0.14\,, n = 6$ ومن الجداول المناظرة لتوزيع t نجد أنّ $t_{\alpha/2} = t_{0.01} = 3.365$ عندما $\nu = n - 1 = 5$. وعليه باستخدام العلاقة (5.11) أعلاه نجد أنّ أعلى خطأ ممكن هو: $\varepsilon = t_{\alpha/2}.\frac{S}{\sqrt{n}} = 3.365.\frac{0.14}{\sqrt{6}} = 0.19$ وهذا يعني بأنّ الخطأ في تقدير المتوسط الحسابي للمجتمع الإحصائي سيكون على الأكثر بمقدار 0.19 وبدرجة ثقة 98٪.

ومنها نجد أنّ التقدير النقطي للمتوسط الحسابي لانصهار المادة المعدنية هو:

$$\mu = \bar{x} \pm \varepsilon = \bar{x} \pm 0.19 = 232.26 \pm 0.19$$

أي أنّ: $\mu_2 = 232.07\,, \mu_1 = 232.45$

3-3-5 تقدير المتوسط الحسابي النقطي μ عندما يكون المجتمع الإحصائي محدوداً أو متناهياً

إذا كان المجتمع الإحصائي قيد الدراسة محدوداً أو متناهياً وكان حجم العينة هو $n \geq 30$ عنصراً فإنّ ذلك يعني بأنّ التوزيع التكراري للمعلمة المراد تقديرها يتقارب مع التوزيع الطبيعي $N(\mu, \sigma^2)$ وعليه نستخدم نفس العلاقات السابقة مع بعض التعديل على قانون حساب أكبر خطأ للتقدير كالتالي:

$$\varepsilon = \bar{x} \quad \mu - z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N\ 1}} \qquad (5.12)$$

حيث N: حجم المجتمع الإحصائي المحدود

n: حجم العينة المأخوذة

ومن العلاقة السابقة يمكننا حساب التقدير النقطي للمعلمة μ حيث نجد:

$$\mu = \bar{x} \pm z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad (5.13)$$

وعندما يكون حجم العينة $n < 30$ فإننا نستنتج بأنّ التوزيع التكراري لمعلمة المجتمع الإحصائي لا يتقارب من التوزيع الطبيعي $N(\mu,\sigma^2)$ وعليه نستخدم بدلاً من ذلك توزيع ستيودنت **t** إذا كان توزيع المجتمع طبيعياً أي الاستعاضة بـ $t_{\alpha/2}$ بدلاً من $z_{\alpha/2}$ في كل من العلاقتين (5.12) و (5.13) لتصبحا كالتالي:

$$\varepsilon = \bar{x} - \mu = t_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \qquad \textbf{(5.14)}$$

$$\mu = \bar{x} \pm t_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \qquad \textbf{(5.15)}$$

وبدرجة حرية $v = n - 1$. وتجدر الإشارة إلى إمكانية استخدام انحراف العينة S بدلاً من σ عندما لا يكون انحراف المجتمع الإحصائي معروفاً. وعندما يكون توزيع المجتمع الأم غير طبيعي فترفض التجربة الإحصائية.

5-3-4 التقدير المجالي لمتوسط المجتمع الإحصائي μ :

لاحظنا مما سبق بأن التقدير النقطي للوسيط μ ينحصر بحساب قيمة أعلى خطأ ε في التقدير. وكثيراً ما تختلف قيمة الوسيط μ عن ذلك وعلى هذا الأساس يفضل في كثير من الأحيان استبدال التقدير النقطي للوسيط μ بالتقدير المجالي أو كما يصطلح عليه بفترة الثقة (confidence interval) حيث يمكننا التأكيد وبدرجة معقولة من الثقة أو بتعبير آخر بدرجة احتمال عالية P بأنّ الوسيط μ يقع ضمن فترة التقديرات المحسوبة وقد لاحظنا في هذا الصدد بأنّ مقدار الخطأ النسبي ε يقع ضمن الفترة التالية:

$$-Z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < Z_{\alpha/2} \qquad \textbf{(5.16)}$$

وذلك عندما يكون تباين المجتمع σ^2 معلوماً حيث درجة الثقة تكون $p = (1-\alpha)\,\%$ وبترتيب العلاقة أعلاه نجد أنّ مجال الثقة لمتوسط المجتمع μ يأخذ الشكل التالي:

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \qquad \textbf{(5.17)}$$

ونطلق على المتراجحة الأخيرة أعلاه أسم مجال الثقة الطبيعي لتقدير متوسط المجتمع الإحصائي μ (normal confidence interval).

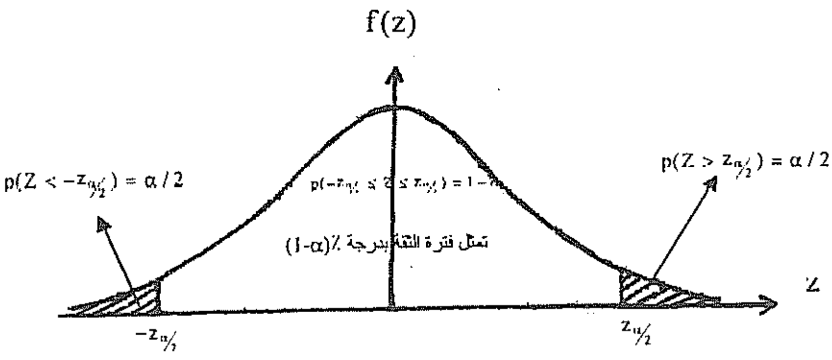
وتبقى نفس المتراجحة (5.17) هي المستخدمة في الحالة التي يكون فيها تباين المجتمع σ^2 مجهولاً وحجم العينة كبيراً $(n \geq 30)$ حيث نَستبدل في العلاقة (5.17) تباين المجتمع الإحصائي σ^2 بتباين العينة العشوائية S^2. أمّا في الحالة التي يكون فيها σ^2 مجهولاً وحجم العينة $(n < 30)$ والمجتمع الأم طبيعي فإننا نجد من نظرية التوزيعات التكرارية للعينات بأنّ الخطأ النسبي ε في هذه الحالة يقع ضمن الفترة التالية:

$$-t_{\alpha/2} < \frac{\bar{x}-\mu}{S/\sqrt{n}} < +t_{\alpha/2} \qquad \textbf{(5.18)}$$

وبدرجة ثقة $p = (1-\alpha)\ \%$ ودرجة حرية $v = n - 1$. ومن العلاقة أعلاه نجد أنّ التقدير المجالي للوسيط μ يأخذ الشكل التالي:

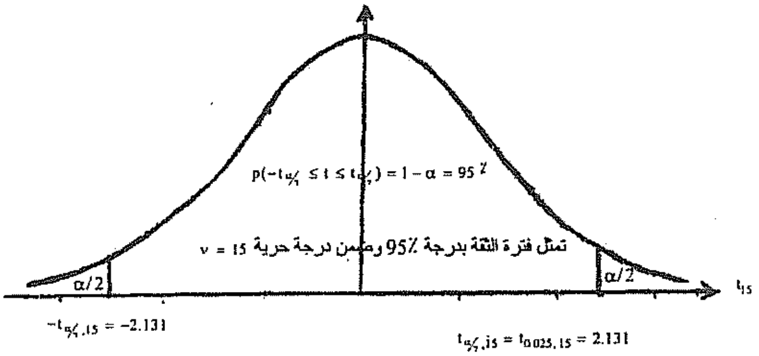
$$\bar{x} - t_{\alpha/2}.\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}.\frac{s}{\sqrt{n}} \qquad \textbf{(5.19)}$$

ونطلق على المتراجحة الأخيرة أسم مجال أو فترة (t) ستيودنت لتقدير μ.
(student-t confidence interval)



الشكل (5-1): يوضح فترتي الثقة وعدم الثقة في التوزيع الطبيعي المعياري

$f(t)$



الشكل (2-5): يوضح فترتي الثقة وعدم الثقة في التوزيع التكراري (ستيودنت -t) حيث درجات الحرية $v = 15$ واحتمال الرفض $\alpha = 0.05$

مثال (11): استخدمت عينة عشوائية من مائة محرك لدراسة قدرة مجتمع إحصائي من المحركات ذي انحراف معياري $\sigma = 5.1$ حصان فإذا كان المتوسط الحسابي للعينة المأخوذة من المجتمع هو $\overline{x} = 21.6$ حصان فأوجد فترة الثقة لقيم μ المحتملة وبدرجة ثقة 95٪.

الحل: بتطبيق متراجحة الثقة الطبيعية ولأن القدرة تتوزع طبيعياً (أي متماثلة) حول متوسطها نجد أنّ:

$$\overline{x} - z_{\alpha/2} . \frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$21.6 - 1.96 . \frac{5.1}{\sqrt{100}} < \mu < 21.6 + 1.96 (\frac{5.1}{\sqrt{100}})$$

أو

$$20.6 < \mu < 22.6$$

وهذا يعني أننا على ثقة بدرجة 95٪ بأنّ الفترة أعلاه تحتوي على قيمة μ أو أننا على ثقة بدرجة 95٪ بأن μ لا تتعدى حدّيها الأعلى $U = 22.6$ والأدنى $L = 20.6$

مثال (12): إذا كان متوسط الوزن المفقود لعينة من أحجار الفرامل متكونة من 16 حجراً بعد مسير 200 كيلو متر هو $\overline{x} = 3.42$ غرام وكان الانحراف المعياري لعناصر العينة هو 0.68 غرام فما هو تقدير المتوسط الحسابي للوزن المفقود للمجتمع الإحصائي (أحجار الفرامل) وبدرجة ثقة 99٪.

الحل: نلاحظ بأنّ عدد عناصر العينة هو $n = 16 < 30$ وأنّ الانحراف المعياري (أو التباين) للمجتمع الإحصائي غير معروف وأنّ توزيع الوزن المفقود طبيعي لذا نستخدم متباينة **t** لتحديد تقدير متوسط المجتمع μ كالتالي:

$$\bar{x} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

$$3.42 - t_{0.005} \cdot \frac{0.68}{\sqrt{16}} < \mu < 3.42 + t_{0.005} \cdot \frac{0.68}{\sqrt{16}}$$

ومنها نجد أنّ: $2.92 < \mu < 3.92$

حيث: $v = 15 \text{ , } t_{0.005} = 2.947 \text{ , } \alpha = 0.01$ وهذا يعني وجوب تغيير أحجار الفرامل بعد مسير 20000 كيلومتر لأن الحد الأدنى للوزن المفقود منها هو 2.92 غراماً وهو ما يعني تلف أحجار الفرامل كلياً وبدرجة ثقة 99٪.

مثال (13): إذا أُخذت عينة عشوائية من الدرجات الامتحانية حجمها $n = 50$ درجة من 200 درجة في امتحان لطلاب هندسة الحواسيب، وكان متوسطها الحسابي $\bar{x} = 75$ وانحرافها المعياري $S = 10$ فالمطلوب:

أ- أوجد فترة الثقة لتقدير المتوسط الحسابي μ للمجتمع الإحصائي لدرجات الطلبة وبدرجة ثقة 95٪.

ب- حدّد درجة الثقة في فترة التقدير التالية: $74 < \mu < 76$

الحل: أ- نلاحظ بأنّ حجم المجتمع الإحصائي محدود وصغير $N = 200$ بالمقارنة مع حجم العينة $n = 50$ لذا نستخدم المتباينة الطبيعية (المحدودة أو المصححة) للتقدير التالية:

$$\bar{x} - z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} < \mu < \bar{x} - z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

حيث استبدلنا الانحراف المعياري المجهول σ بالانحراف المعياري S.

وبالتعويض نجد أنّ: $72.6 < \mu < 77.4$

ب- نلاحظ باستخدام نفس متباينة الثقة أعلاه $74 < \mu < 76$ أنّ:

$$74 = 75 - z_{\alpha/2} \cdot \frac{10}{\sqrt{50}} \sqrt{\frac{150}{199}} < \mu < 75 + z_{\alpha/2} \cdot \frac{10}{\sqrt{50}} \sqrt{\frac{200-50}{199}} = 76$$

ومنها مقدار الخطأ هو: $1 = \varepsilon = z_{\alpha/2} \cdot \frac{10}{\sqrt{50}} \sqrt{\frac{150}{199}}$

أي أنّ: $1.23\ z_{\alpha/2} = 1$

ومنه: $z_{\alpha/2} = \frac{1}{1.23} = 0.81$.

وباستخدام جدول المنحني الطبيعي المعياري نجد أنّ المساحة تحت المنحني من $z = 0$ إلى $z = 0.81$ هي: $P = \left(\frac{1-\alpha}{2}\right) = 0.2910$ أي أنّ:

$P = 1 - \alpha = 2(0.2910) = 0.582$ وبالتالي فإنّ درجة الثقة المطلوبة هي: $P = 58.2\%$

5-3-5 تقدير الفرق $\mu_1 - \mu_2$ (أو المجموع) بين متوسطي مجتمعين إحصائيين معلومي التباين وعندما يكون حجم العينات كبيراً:

في كثير من الحالات يود الباحث معرفة الفوارق بين مجتمعين إحصائيين خصوصاً في المنتجات الدوائية أوالمنتجات الزراعية أوالمنتجات الصناعية وبديهي عندما تكون الفوارق ليست بذات أهمية أو تكاد تنعدم فإنّ ذلك يعني تكافؤ المنتجات وعلى هذا الأساس يُبنى قرار شراء مثل هذه المنتجات.

فلو افترضنا بأن لدينا مجتمعين إحصائيين صناعيين ذوا متوسطين μ_2 ، μ_1 وتباينين σ_1^2 ، σ_2^2 على الترتيب فإنّ أفضل وسيلة لدراسة الفوارق بين ا لمجتمعين هو التقدير للفرق بين الوسيطين μ_1 و μ_2 أي التقدير للمقدار $\mu_1 - \mu_2$ وذلك باستخدام الإحصائية $\bar{x}_D = \bar{x}_1 - \bar{x}_2$ حيث $\bar{x}_1$ تمثل متوسط عينة ذات حجم n_1 من المجتمع الأول.

$\bar{x}_2$ تمثل متوسط عينة ذات حجم n_2 من المجتمع الثاني.

ولتوضيح عدم انحياز المقدّر الجديد $\bar{x}_D$ في تقدير الفرق $\mu_1 - \mu_2$ في حالة كون المتغير العشوائي $\bar{X}_i$ منقطعاً نلاحظ بأنّ: $B = E(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2) = 0$ لأنّ:

$$E(\bar{X}_D) = E(\bar{X}_1 - \bar{X}_2) = \frac{1}{n}E\left[\sum_{i=1}^{n}(X_{1i} - X_{2i})\right] = \frac{1}{n}\left[\sum_{i=1}^{n}E(X_{1i} - X_{2i})\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[E(X_{1i}) - E(X_{2i})\right] = \frac{1}{n}(n\mu_1 - n\mu_2) = \mu_1 - \mu_2$$

حيث أخذنا العينتين متساويتي الحجمين $(n_1 = n_2)$. أمّا عدم انحياز المعلمة $\bar{X}_D$ في الحالة المتصلة أي عندما يكون المتغير العشوائي X_i متصلاً فيتوضح بالتالي:

$$E(\overline{X}_D)=E(\overline{X}_1-\overline{X}_2)=\int_{-\infty}^{\infty}......\int_{-\infty}^{\infty}(\overline{x}_1-\overline{x}_2)f(x_{11}.....x_{1n_1},x_{21}...x_{2n_2})\,d\overline{x}_1 d\overline{x}_2$$

وبما أن المتغيرات X_{ij} مستقلة لذا فإنّ دالة الكثافة قابلة للتجزئة الضربية أي أنّ:

$$E(\overline{X}_1-\overline{X}_2)=\int_{-\infty}^{\infty}......\int_{-\infty}^{\infty}\overline{x}_1 f(x_{11}......x_{1n_1})\,dx_{11}...dx_{1n_1}$$
$$-\int_{-\infty}^{\infty}......\int_{-\infty}^{\infty}\overline{x}_2\ \ f(x_{21}......x_{2n_2})\ \ dx_{21}...dx_{2n_2}$$

ومنه نجد أن:
$$E(\overline{X}_D)=E(\overline{X}_1-\overline{X}_2)=\mu_1-\mu_2$$

وبناءً على نظرية التوزيع الطبيعي للمتوسطات فإنه يمكن اعتبار التوزيع التكراري لفرق المتوسطين $\overline{X}_D=\overline{X}_1-\overline{X}_2$ توزيعاً طبيعياً أيضاً كما يمكن استخدام متراجحة الثقة الطبيعية وبدرجة ثقة أو احتمالية ٪ $p=(1-\alpha)$ علماً بأنّ العدد الطبيعي للفروق أو الدرجة المعيارية للفروق تحسب من العلاقة التالية:

$$z=\frac{\overline{x}_D-\mu_D}{\sigma_D}=\frac{(\overline{x}_1-\overline{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \qquad \textbf{(5.20)}$$

يقع ضمن مجال الثقة الطبيعي: $-z_{\alpha/2}<z<z_{\alpha/2}$

ومنها نجد أنّ متراجحة الفرق الطبيعية للمعلمتين $\mu_1-\mu_2$ تأخذ الصورة التالية:

$$(\overline{x}_1-\overline{x}_2)-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}<\mu_1-\mu_2<(\overline{x}_1-\overline{x}_2)+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \qquad \textbf{(5.21)}$$

ومن مجال الثقة أعلاه يمكننا الاستفادة للحصول على التقدير النقطي لوسيط الفرق $\mu_D=\mu_1-\mu_2$ وذلك باستخدام الحدين الأعلى أو الأدنى كتقديرين مختلفين لها:

$$\mu_D=\mu_1-\mu_2\cong U=\overline{x}_1-\overline{x}_2+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \qquad \textbf{(5.22)}$$

$$\mu_D=\mu_1-\mu_2\cong L=\overline{x}_1-\overline{x}_2-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \qquad \textbf{(5.23)}$$

ملاحظة ③: إذا كان تبايني المجتمعين الإحصائيين σ_1^2 و σ_2^2 مجهولاً وكان حجم العينتين المأخوذتين من هذين المجتمعين كبيراً بقدر كاف (أي $n_1,n_2\geq 30$) عندئذ يبقى التوزيع

المستخدم هو التوزيع الطبيعي المعياري وتبقى العلاقات (5.20)، (5.21)، (5.22)، (5.23) هي المستخدمة مع تبديل σ_1^2, σ_2^2 في هذه العلاقات بـ S_1^2, S_2^2.

أمّا إذا كان σ_1^2 و σ_2^2 مجهولاً وكان حجما العينتين صغيراً (أي $n_1, n_2 < 30$) فإننا سنوضح هذه الحالة من خلال الفقرة القادمة.

5-3-6 تقدير الفرق $\mu_1 - \mu_2$ (أو المجموع) بين متوسطي مجتمعين إحصائيين غير معلومي التباين وعندما يكون حجم العينات صغيراً:

عندما يكون لدينا مجتمعين إحصائيين غير معلومي التباين ولكن توزيعهما التكراري من النوع الطبيعي أو المعتدل وعندما يكون حجم كل من العينتين المأخوذتين من هذين المجتمعين صغيراً ($n_1 < 30, n_2 < 30$) فإننا نستخدم عادةً متباينة الثقة الخاصة بتوزيع t التالية بدلاً من توزيع Z. أيّ:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \qquad \textbf{(5.24)}$$

حيث t حسب نظرية التوزيع التكراري تساوي: $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - \mu_D}{\hat{\sigma}^2_{\bar{x}_1 - \bar{x}_2}}$

حيث $S^2 = \hat{\sigma}^2_{\bar{x}_1 - \bar{x}_2}$ هو مقدّر التباين. وقد نفترض أحياناً للسهولة تساوي التباينات أيّ:

$\sigma^2 = \sigma_1^2 = \sigma_2^2$ ولتقدير S^2 نستخدم ما يسمى بالتباين التجميعي (أو سلة التباينات) (pooling variance)

$$\hat{\sigma}^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \cong \sigma_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \qquad \textbf{(5.25)}$$

$$S_p^2 = \hat{\sigma}_p^2 = \frac{\sum(x_{1i} - \bar{x}_1) + \sum(x_{2i} - \bar{x}_2)}{(n_1 - 1) + (n_2 - 1)} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \qquad \textbf{(5.26)}$$

حيث هنالك $n_1 - 1$ من الانحرافات (ودرجات الحرية) في العينة الأولى و $n_2 - 1$ من الانحرافات (أودرجات الحرية) في العينة الثانية. وبتعويض (5.26) في (5.25) ومن ثم في العلاقة t السابقة نجد أنّ:

$$t=\frac{(\bar{x}_1-\bar{x}_2)-\mu_D}{\sqrt{(n_1-1)S_1^2+(n_2-1)S_2^2}}\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}} \qquad (5.27)$$

وهذا المتغير العشوائي ذو التوزيع التكراري (ستيودنت -t) يقترب من التوزيع الطبيعي بدرجة من الحرية هي: $v=n_1+n_2-2$ ويقع ضمن متباينة الثقة التالية وذلك بدرجة ثقة $p=(1-\alpha)\%$.

$$-t_{\alpha/2}<t<t_{\alpha/2}$$

أو:

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}S_p\sqrt{\frac{n_1+n_2}{n_1n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\alpha/2}S_p\sqrt{\frac{n_1+n_2}{n_1n_2}} \qquad (5.28)$$

حيث افترضنا $\sigma_2^2=\sigma_1^2$ وغير معروفين.

ملاحظة ④: في بعض الحالات نجد أنّ مجموع حجمي العينتين يكون كبيراً على رغم صغر حجميهما أي أنّ: $n_1+n_2>30$ مثلاً وفي مثل هذه الحالات نستخدم القيم الحرجة التالية:

الجدول (5-1)

v	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.025$	$\alpha=0.01$	$\alpha=0.005$
(inf > 30)	1.282	1.645	1.960	2.326	2.576

حيث يرمز inf إلى أكبر حد أدنى للمجموع.

ملاحظة ⑤: عند افتراض عدم تساوي التباينين $\sigma_2^2\neq\sigma_1^2$ وعدم معرفتهما ولتحديد فترة الثقة أو فترة التقدير لفرق المتوسطين $\mu_D=\mu_1-\mu_2$ للعينات صغيرة الحجم $n_1\neq n_2$ والمأخوذة من مجتمعين إحصائيين مختلفين فإننا نستخدم متباينة الثقة أو فترة التقدير التالية الخاصة بتوزيع t والمناظرة للعلاقة السابقة:

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\mu_1-\mu_2<x_D+t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \qquad (5.29)$$

حيث $t_{\alpha/2}$ هي القيمة للدرجة المعيارية الثائية بدرجة ثقة $p=1-\alpha$ ودرجة حرية:

$$v=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\left[\left(\frac{S_1^2}{n_1}\right)^2\Big/(n_1-1)\right]+\left[\left(\frac{S_2^2}{n_2}\right)^2\Big/(n_2-1)\right]} \tag{5.30}$$

مثال (14): إذا كان متوسط تسرب الزيت الخفيف لعينة من 15 محرك سيارة كيا هو 4.93 ملم٣ وبانحراف معياري 1.14 ملم٣ بينما كان تسرب الزيت الخفيف لعينة من عشرة محركات من سيارات هونداي هو 2.64 ملم٣ وبانحراف معياري 0.66 أحسب فترة الثقة أو التقدير لفرق متوسطي تسريب الزيت الخفيف لكلا المجتمعين من المحركات كيا وهونداي معتبراً درجة الثقة أو الاحتمالية هي $p = 1\text{-}\alpha = 95\%$ علماً بأنّ تبايني المجتمعين مختلفين أيضاً وغير معروفين.

الحل: من السؤال أعلاه نجد المعلومات التالية:

$\bar{x}_1 = 4.93$	$\bar{x}_2 = 2.64$
$S_1 = 1.14$	$S_2 = 0.66$
$n_1 = 15$	$n_2 = 10$

$\alpha = 0.05$ وأنّ $\sigma_1^2 \neq \sigma_2^2$ وغير معلومين.

والمطلوب هو حساب فترة التقدير (الثقة) للمعلمة $\mu_D = \mu_1 - \mu_2$ وباحتمالية أو درجة ثقة 95٪ .

وبما أنّ التباينين غير متساويين وحجم العينتين غير متساوٍ أيضاً لذا نستخدم متباينة الثقة الخاصة بتوزيع t وبدرجة حرية هي:

$$v=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2\Big/(n_1-1)+\left(\frac{S_2^2}{n_2}\right)^2\Big/(n_2-1)}=\frac{\left(\frac{(1.14)^2}{15}+\frac{(0.66)^2}{10}\right)^2}{\left(\frac{(1.14)^2}{15}\right)^2\Big/14+\left(\frac{(0.66)^2}{10}\right)^2\Big/9}$$

$$=22.7 \cong 23$$

وبما أنّ $\alpha = 0.05$ لذا فإنّ $t_{\alpha/2} = t_{0.025} = 2.069$

وبالتعويض في العلاقة (5.29) السابقة نجد أنّ:

$$2.29 - 2.069\sqrt{\frac{(1.14)^2}{15} + \frac{(0.66)^2}{10}} < \mu_1 - \mu_2 < 2.29 + 2.069\sqrt{\frac{(1.14)^2}{15} + \frac{(0.66)^2}{10}}$$

أو أنّ: $2.02 < \mu_1 - \mu_2 < 2.26$

وعلى هذا الأساس نجد أنّ الفارق بين المجتمعين وبدرجة ثقة 95٪ ليس كبيراً وهو يكاد يصل إلى 0.24 ومن هذا الفرق الضئيل نستطيع الجزم وبمقدار 95٪ بأن المحركات تكـــاد تكـــون متساوية أو متكافئة في التسريب للزيت الخفيف.

ملاحظة ⑥: لاحظ أنّ كافة العمليات السابقة ونقصد اختيار العينات كان بصورة عشوائية غير منحازة ومستقلة غير مترابطة مع بعضها أمّا حالة عدم العشوائية أو عدم الاستقلالية فإنّ لهـــا حسابات مختلفة ولكنها مشابهة لفترات التقدير السابقة.

مثال (15): تدربت مجموعتان من المهندسين مكونتان من 12 ، 10 مهندسين على الترتيب على برنامج لصيانة محركات الديزل في ورشتين مختلفتين B,A فإذا كان متوسط الدرجات للمجموعة الأولى هو 85 وبانحراف معياري 4 درجة بينما كان متوسط الدرجات للمجموعة الثانية هو 81 وبانحراف معياري 5 درجة أوجد فترة الثقة أو التقدير للفرق بين المتوسطين وبدرجة ثقة 90٪ على افتراض أنّ المجتمعين الإحصائيين معتدلا التوزيع وذوا تباين متساو.

الحل: فرق المتوسطات، $\bar{x}_D = \bar{x}_1 - \bar{x}_2 = 85 - 81 = 4$ نلاحظ أنّ: $n_1 < 30$, $n_2 < 30$ وكذلك $\sigma_1^2 = \sigma_2^2$ وغير معروفين وعليه نستخدم تقدير التباين التجميعي:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(11)(16) - (9)(25)}{12 + 10 - 2} = 20.05$$

ومنه فإنّ: $S_p = 4.478$ وبما أنّ $p = 1 - \alpha = 0.90$ فإنّ $\alpha = 0.10$ حيث: $v = n_1 + n_2 - 2 = 20$ وبما أنّ حجم العينتان صغير $n_1, n_2 < 30$ لذا نستخدم فترة أو متباينة التقدير الخاصة بتوزيع ستيودنت **t** حيث نجد من الجداول المناظرة أنّ: $t_{\alpha/2} = t_{0.05} = 1.725$ وعليه فإنّ:

$$\bar{x}_D - t_{\alpha/2} S_p \sqrt{\frac{n_1 + n_2}{n_1 n_2}} < \mu_1 - \mu_2 < \bar{x}_D + t_{\alpha/2} S_p \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

ومنها نجد أنّ: $0.69 < \mu_D < 7.31$ وعليه يمكن القول بأننا على ثقة بمقدار 90٪ من أنّ الفارق الحقيقي بين المتوسطين كبير نوعاً ما ويقع ضمن الفترة السابقة كما أنّ النتيجة الموجبة للفرق بين المتوسطين تدُل على أنّ برنامج الصيانة في الورشة A أفضل منه في الورشة B.

5-4 تقدير تباين المجتمع الإحصائي σ^2:

تعتبر أفضل وسيلة لتقدير تباين المجتمع الإحصائي σ^2 هي استخدام تباين العينة العشوائية غير المنحازة S^2، وبالمثل تقدير الانحراف المعياري له σ بالانحراف المعياري S للعينة وذلك لأنّ S^2 معلمة غير منحازة أيضاً أي أنّ: $B = E(S^2) - \sigma^2 = 0$ حيث:

ولتوضيح خاصية عدم الانحياز في الحالة المتصلة (عندما X_i متغيـر عشوائي متصل) نلاحظ بأنّ: $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

$$E(S^2) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} S^2 f(x_1, x_2 \cdots x_n) dx_1 \cdots dx_n$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{\sum(x_i - \bar{x})^2}{n-1} f(x_1, \cdots x_n) dx_1 \cdots dx_n$$

$$= \frac{1}{n-1}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (\sum x_i^2 - n\bar{x}^2) f(x_1, \cdots x_n) dx_1 \cdots dx_n$$

$$= \frac{1}{n-1}\sum\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_i^2 f(x_1, \cdots x_n) dx_1 \cdots dx_n - \frac{n}{n-1}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \bar{x}^2 f(\bar{x}) d\bar{x}_i$$

وبتحليل كل من التكاملين بصورة منفردة نجد أنّ التكامل الأول يؤول إلى:

$\frac{1}{n-1}\sum_{i=1}^{n}\int_{-\infty}^{\infty} x_i^2 f(x_i) dx_i = \frac{n}{n-1}\sigma^2$ حيث البيانات x_i مستقلة عن بعضها البعض.

وحيث $\int_{-\infty}^{\infty} f(x_j) dx_j = 1$ عندما $i \neq j$

أمّا التكامل الثاني فيساوي:

$$\frac{n}{n-1}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \bar{x}^2 f(x_1, \cdots x_n) dx_1 \cdots dx_n$$

$$= \frac{n}{n-1}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \bar{x}^2 f(x_1)\cdots f(x_n)\ dx_1 \cdots dx_n$$

$$= \frac{n}{n-1}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{(\sum xi)^2}{n^2} f(x_1)\cdots f(x_n)\ dx_1 \cdots dx_n$$

$$= \frac{n}{n^2(n-1)}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (\sum x_i^2 + 2\sum\sum_{i \neq j} x_i x_j) f(x_1) \ldots f(x_n) dx_1 \cdots dx_n$$

$$= \frac{n}{n^2(n-1)}\sum\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_i^2 f(x_1) \ldots f(x_n) dx_1 \cdots dx_n$$

$$+ \frac{2n}{n^2(n-1)}\sum\sum\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (x_i x_j) f(x_1) \ldots f(x_n) dx_1 \cdots dx_n$$

$$=\frac{1}{n(n-1)}\sum_{i=1}^{n}\int_{-\infty}^{\infty}x_i^2 f(x_i)dx_i+\frac{2}{n(n-1)}\sum\sum\int_{-\infty}^{\infty}x_i f(x_i)dx_i\int_{-\infty}^{\infty}x_j f(x_j)dx_j$$

$$=\frac{1}{n(n-1)}n\sigma^2=\sigma^2/(n-1)$$

حيث $\sum_{i=1}^{n}\int_{-\infty}^{\infty}x_i^2 f(x_i)dx_i=\sigma^2$

وحيث $\int_{-\infty}^{\infty}x_i f(x_i)dx_i=\int_{-\infty}^{\infty}x_j f(x_j)dx_j=0$

وبالجمع نحصل على التالي: $E(S^2)=\frac{n}{n-1}\sigma^2-\frac{\sigma^2}{n-1}=\frac{n-1}{n-1}\sigma^2=\sigma^2$

وكما لاحظنا في التقدير النقطي (P.E) لمتوسط المجتمع الإحصائي μ نلاحظ أيضاً بأنّ التقدير النقطي لتباين المجتمع الاحصائي يتمثل بتحديد التقدير بحدي متباينة الثقة الأعلى (U) والأدنى (L) أو أي قيمة مفردة بينهما.

5-4-1 التقدير المجالي للوسيط σ^2:

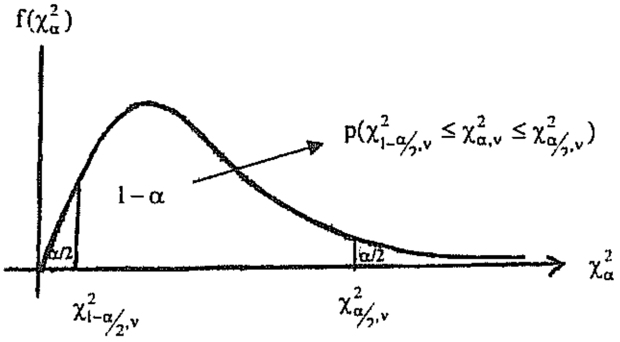
لتقدير قيمة تباين المجتمع الإحصائي σ^2 والذي يتميز بتوزيعه الطبيعي يمكن الاستفادة من تعريف المتغير العشوائي χ^2_α والذي يفترض أن تكون العينة العشوائية مأخوذة من مجتمع إحصائي ذي توزيع طبيعي.

$$\chi^2_\alpha=\frac{(n-1)\,S^2}{\sigma^2} \qquad \textbf{(5.31)}$$

حيث تمثل القيمة n حجم العينة، S^2 تمثل تباين العينة العشوائية صغيرة كانت أم كبيرة، σ^2 تمثل تباين المجتمع الإحصائي.

ومنها نجد أنّ متباينة الثقة أو التقدير المجالي للوسيط σ^2 وبدرجة احتمال $p=1-\alpha$ يُعطى بالشكل التالي:

$$\chi^2_{1-\alpha/2,\nu}<\frac{(n-1)\,S^2}{\sigma^2}<\chi^2_{\alpha/2,\nu} \qquad \textbf{(5.32)}$$



الشكل (5-3)

وبإعادة ترتيب المتراجحة أعلاه نجد التالي:

$$\frac{(n-1)\,S^2}{\chi^2_{\alpha/2,v}} \le \sigma^2 \le \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2}} \qquad \textbf{(5.33)}$$

ويمكن الاستفادة من المتباينة أعلاه في تقدير الانحراف المعياري σ للمجتمع الإحصائي. وعندما يكون حجم العينة كبيراً $n \ge 30$ فإنه من الممكن اعتبار توزيع العينات في المجتمع الإحصائي توزيعاً يقترب كثيراً من التوزيع الطبيعي ذي المتوسط الحسابي المساوي للانحراف المعياري للمجتمع $\mu = \sigma$ والانحراف المعياري الذي يساوي $\frac{\sigma}{\sqrt{n}}$ وعلى هذا الأساس تحسب الدرجة المعيارية z والتي تمثل متغيراً عشوائياً بالعلاقة التالية:

$$z = \frac{S-\sigma}{\sigma/\sqrt{n}} \qquad \textbf{(5.34)}$$

والذي يقع ضمن المتباينة الطبيعية التالية:

$$-z_{\alpha/2} < \frac{S-\sigma}{\sigma/\sqrt{n}} < z_{\alpha/2} \qquad \textbf{(5.35)}$$

وبإعادة ترتيب المتراجحة السابقة نحصل على فترة الثقة الطبيعية لتقدير σ

$$\frac{S}{1+z_{\alpha/2}/\sqrt{n}} < \sigma < \frac{S}{1-z_{\alpha/2}/\sqrt{n}} \qquad \textbf{(5.36)}$$

حيث نستخدم المتراجحة أعلاه للعينات كبيرة الحجم $n \geq 30$ وبدرجة ثقة أو احتمالية $p = (1-\alpha)$ ٪ وفيما يلي نستعرض بعض الأمثلة لتوضيح كيفية الاستفادة من العلاقات السابقة:

مثال (16): يستورد أحد المصانع عدد من الألواح الزجاجية لصناعة العدسات ومن خبرة المصنع السابقة تم تحديد تباين معامل الانكسار لهذا النوع من الألواح الزجاجية بالقيمة $\sigma^2 = 1.26 \times 10^{-4}$ ولما كان من الأهمية بمكان أن تكون كافة الألواح الزجاجية من نفس معامل الانكسار فإنّ المصنع يرفض الشحنة المستوردة إذا كان تباين أي عينة عشوائية ذات حجم $n = 20$ لوحة زجاجية هو $S^2 = 2.00 \times 10^{-4}$ ولو افترضنا بأن مجتمع الألواح الزجاجية هو مجتمع طبيعي التوزيع:

أ- فما هو احتمال رفض الشحنة المستوردة حتى لو كان $\sigma^2 = 1.26 \times 10^{-4}$.

ب- إذا افترضنا بأن تباين معامل الانكسار للألواح الزجاجية في عينة عشوائية ذات حجم $n = 20$ هو $S^2 = 1.20 \times 10^{-4}$ فاحسب فترة (تقدير) الثقة للانحراف المعياري للمجتمع الإحصائي وبدرجة احتمال (ثقة) 95٪.

الحل: أ- نلاحظ بأنّ الكمية المستوردة هي كمية محدودة أي أنّ المجتمع الإحصائي من النوع المحدود والمتناهي وأن حجم العينة ليس كبيراً لذا نستخدم قاعدة المتغير العشوائي لمربع كاي χ^2 المبينة بالعلاقة (5.31):

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2} = \frac{19(2.00 \times 10^{-4})}{1.26 \times 10^{-4}} = 30.2$$

وباستخدام جداول مربع كاي χ^2_α وبدرجة حرية $v = n - 1 = 19$ نجد أنّ الاحتمال المقابل للعدد المذكور $\chi^2_\alpha = 30.2$ هو: $\alpha = 0.05$ أي أنّ احتمال رفض الشحنة هو $q = 0.05$ ومنها نجد أيضاً أنّ احتمال القبول للشحنة هو $P = 1 - \alpha = 0.95$.

ب- من درجة الثقة المذكورة 95٪ نجد أنّ $\alpha = 0.05$ ، $\frac{\alpha}{2} = 0.025$ وأنّ $1 - \frac{\alpha}{2} = 0.975$

كما نلاحظ بأنّ الأعداد المقابلة لهذه الاحتمالات في توزيع مربع كاي هي كالتالي:

$\chi^2_{\alpha/2} = \chi^2_{0.025} = 32.852$ ، $\chi^2_{1-\alpha/2} = \chi^2_{0.975} = 8.907$ وبتطبيق متباينة الثقة الكائية (الخاصة بمربع كاي) نجد التالي:

$$\frac{19(1.20\times10^{-4})}{32.852} < \sigma^2 < \frac{19(1.20\times10^{-4})}{8.907}$$

$$0.000069 < \sigma^2 < 0.000256$$

$$0.69\times10^{-4} < \sigma^2 < 2.56\times10^{-4}$$

ومنه:

$$0.0083 < \sigma < 0.016$$

ويلاحظ من متراجحة التباين أنّ الحد الأعلى هو $U = 2.56\times10^{-4}$ وهو ما يؤكد إمكانية رفض الشحنات لتجاوزها شرط القبول وهو 2.00×10^{-4} ومن البديهي أن نستنتج قبول كل الشحنات لو كان الحد الأعلى لتباين المجتمع الإحصائي هو أقل من شرط القبول السابق كأن يكون مثلاً: $U = 1.999\times10^{-4}$.

مثال 17: فيما يلي أحجام عشر علب مصنوعة في إحدى شركات التعليب:

46, 46.9, 45.8, 45.9, 46.1, 47.0, 45.8, 46.1, 46.4, 54.2

أوجد متباينة الثقة (التقدير المجالي) وبدرجة احتمال(ثقة) 95٪ لتباين أحجام مجتمع علب التعليب المصنوعة في تلك الشركة معتبراً أنّ الحجوم ذات توزيع تكراري معتدل.

الحل: نلاحظ بأن العينة صغيرة $n < 30$ ومأخوذة من مجتمع إحصائي غير محدود لذا نستخدم متباينة مربع كاي χ^2 التالية: $\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}$ ومن العينة المذكورة أعلاه نجد أنّ: $S^2 = \frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)} = 0.286$ ومن درجة الاحتمال نجد أنّ $\alpha = 0.05$ وأنّ مربع كاي المقابل لهذا الاحتمال وضمن درجة حرية تساوي $v = n-1 = 9$ هو:

$\chi^2_{\alpha/2} = \chi^2_{0.025} = 19.023$ ومن الجداول نجد أيضاً بأنّ: $\chi^2_{1-\alpha/2} = \chi^2_{0.975} = 2.7$

وبالتعويض في المتراجحة السابقة نحصل علـــى التـــالي: $\frac{9(0.286)}{19.023} < \sigma^2 < \frac{9(0.286)}{2.7}$

ومنه: $0.135 < \sigma^2 < 0.953$.

ملاحظة ⑦: من المثالين السابقين (16) و (17) نلاحظ إمكانية اســتخدام الحــدين الأعلــى والأدنى لمتراجحتي الثقة أو أي نقطة بينهما لتقدير قيمة تباين المجتمع الإحصائي σ^2 بالصورة النقطية. أي يمكننا القول في المثال (17) بأنّ التقدير النقطي لـ σ^2 هــو: $\sigma^2 = U = 0.953$ أو $\sigma^2 = L = 0.135$ وفقاً لمتطلبات أو شروط العرض أو شروط التصنيع.

5-5 تقدير النسبة بين تبايني مجتمعين إحصائيين $\frac{\sigma_1^2}{\sigma_2^2}$ ذوا توزيع طبيعي:

تعتبر أفضل وســيلة (راجــع شــروط التقــدير) لتقــدير النســبة بــين تبــايني مجتمعــين إحصائيين σ_2^2, σ_1^2 نقطياً هو استخدام النسبة بين تبايني عينتين عشــوائيين منهمــا أي $\left(\frac{S_1^2}{S_2^2}\right)$ وذلك بفرض أنّ حجم كل من العينتين هو n_2, n_1 ودرجتي الحرية المقابلــة للعينتــين همــا $v_1 = n_1 - 1 \quad , \quad v_2 = n_2 - 1$ على الترتيب أما التقدير المجالي لتلك النسبة فيتم باستخدام إحصائية فيشر (راجع التوزيع التكراري لفيشر) التالية:

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\sigma_2^2 \; S_1^2}{\sigma_1^2 \; S_2^2} \qquad \textbf{(5.37)}$$

والتي تأخذ عيناتها توزيعاً تكرارياً خاصاً يسمى بتوزيع فيشر، ومن ناحية نظريــة نجــد أنّ الإحصائية F أعلاه هي في الحقيقة عبارة عن النسبة بين متغيــرين مســتقلين لمربــع كــاي χ_1^2 ، χ_2^2 وبدرجتي حرية $v_1 = n_1 - 1 \quad , \quad v_2 = n_2 - 1$ على الترتيب أي أنّ:

$$F = \frac{\chi_1^2 / v_1}{\chi_2^2 / v_2}$$

أمّا القيمة الحرجة f_α فتمثل المساحة أو منطقة الرفض بينما $f_{1-\alpha}$ تمثل منطقة القبــول، إنّ فترة التقدير المقبولة هي:

$$f_{1-\alpha/2}(v_1, v_2) < F = \frac{\sigma_2^2\, S_1^2}{\sigma_1^2\, S_2^2} < f_{\alpha/2}(v_1, v_2) \qquad (5.38)$$

حيث: $f_{1-\alpha/2}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}$ وحيث نجد بأنّ احتمال هذه الفترة هو $1\text{-}\alpha$ كما هو المتعارف (راجع نظرية التوزيعات التكرارية) وبالتعويض والترتيب نجد أنّ الفترة أعلاه تكتب كالتالي:

$$\frac{S_1^2}{S_2^2} f_{1-\alpha/2}(v_2, v_1) = \frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2, v_1) \qquad (5.39)$$

وتفيدنا هذه الفترة في استطلاع مدى التقارب بين التباينين فإذا كانت الفترة صغيرة وتشير إلى الواحد مثلاً كان المجتمعان متكافئين وإلاّ فلا.

مثال (18): تم فحص إنتاج شركتين A ,B لصناعة البطاريات الجافة صغيرة الحجم فإذا أُخذت عينة ذات حجم 25 من الشركة A وعينة ذات حجم 16 من الشركة B ووجد أنّ متوسط طول عمر بطاريات العينة الأولى A هو 82 ساعة وبانحراف معياري مقداره 8 ساعة بينما متوسط عمر بطاريات العينة الثانية B هو 78 وبانحراف معياري 7 فأحسب فترة الثقة وبدرجة 98٪ للنسبة بين تبايني إنتاج الشركتين A ,B وعلى افتراض أنّ المجتمعين الإحصائيين ذوا توزيع تكراري طبيعي (معتدل)

الحل: من معلومات السؤال نجد أنّ:

$$v_1 = 24 \quad S_1 = 8 \quad \bar{x}_1 = 82 \quad n_1 = 25$$
$$v_2 = 15 \quad S_2 - 7 \quad \bar{x}_2 = 78 \quad n_2 = 16$$

والمطلوب هو حساب فترة الثقة للنسبة σ_1^2/σ_2^2 وبدرجة ثقة $p = 1 - \alpha = 98\%$ أي أنّ: $\alpha = 0.02$ أو أنّ $\alpha/2 = 0.01$ وباستخدام المتباينة (5.39) السابقة نحصل على التالي:

$$\frac{S_1^2}{S_2^2}\frac{1}{f\alpha/2(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f\alpha/2(v_2, v_1)$$

$$\frac{64}{49}\frac{1}{3.29} = \frac{64}{49}\frac{1}{f_{0.01}(24{,}15)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{64}{49} f_{0.01}(15{,}24) = \frac{64}{49}(2.89)$$

ومنها نجد أنّ فترة الثقة للنسبة بين التباينين هي: $0.397 < \frac{\sigma_1^2}{\sigma_2^2} < 3.775$

ومن المتراجحة أعلاه نلاحظ بأنّ الفارق بين إنتاجي الشركتين كبير جداً وهذا يؤدي إلى عدم تكافؤ إنتاجيهما للبطاريات وبأخذ الجذر التربيعي للمتراجحة نجد أنّ فترة الثقة للنسبة بين انحرافي المجتمعين الإحصائيين: $0.630 < \frac{\sigma_1}{\sigma_2} < 1.943$

5-6 التمثيل الهندسي (البياني) لمجال الثقة:

إذا كانت A تمثل مجال الثقة (التقدير) لمتوسط المجتمع μ، B مجال الثقة (التقــدير) لتبــاين المجتمع σ^2 أي أنّ:

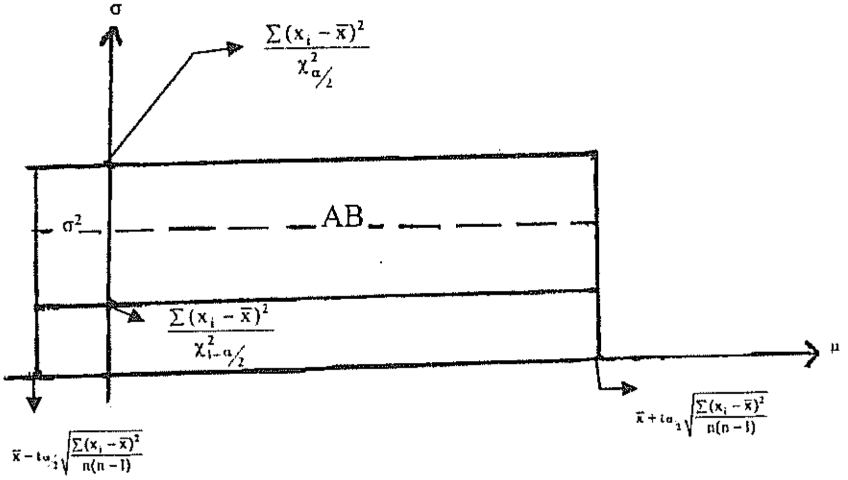
$$P(A) = P\left(L = \bar{x} - t_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2}\frac{S}{\sqrt{n}} = U \right) = 1-\alpha$$

$$P(B) = P\left(L = \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\alpha/2}} = U \right) = 1-\alpha$$

فإنّ احتمال حصول الحدثين B, A شريطة أن يكونا مستقلين هو:

$$P(A\ B) = P(A)\,.\,P(B) = (1-\alpha)^2$$

ومنها يمكننا تحديد منطقة الثقة الجديدة لحصول الحدثين كالتالي وعلى افتراض أنّ σ^2 غيــر معروف:



الشكل (5-4)

وفي حالة معلومية تباين المجتمع الإحصائي σ_0^2 فإنّ المستطيل المذكور A B يؤول إلى خط مستقيم فقط عند النقطة σ_0^2.

5-7 تقدير النسبة P للمجتمع الإحصائي: Estimation of Proportion

تعتبر أفضل وسيلة (راجع شروط التقدير) لتقدير احتمال وقوع حدث معيّن P(e) ضمن مجتمع إحصائي هي استخدام ما يسمى بالتكرار النسبي لوقوع الحدث e ضمن عينة مأخوذة من ذات المجتمع الإحصائي. وبعبارة أخرى إنّ أفضل إحصائية لتقدير احتمال وقوع الحدث e تقديراً نقطياً هو: $\hat{P} = \frac{x}{n}$ حيث x هو تكرار وقوع أو حصول الحدث e خلال عدد n من التجارب الثنائية أو البرنولية المستقلة (independent Bernolli trials). وتُعتبر هذه الإحصائية غير منحازة أساساً لأنّ مقدار الخطأ فيها يساوي صفراً. أي أنّ: $p = 0$ $B = E(\frac{X}{n})$ أو أنّ: $E(\hat{P}) = E(\frac{X_i}{n}) = \frac{1}{n}E(X_i) = \frac{1}{n}(np) = p$ حيث القيمة المتوقعة $E(X_i)$ في التوزيع الثنائي تساوي $\mu = np$. ولإثبات عدم الانحياز في الحالة المتصلة (أي عندما X_i متغير عشوائي متصل) فإنّ:

$$E(\hat{p}) = \int_{-\infty}^{\infty} \frac{x_i}{n} f(x_i)dx_i = \frac{1}{n}\int_{-\infty}^{\infty} x_i f(x_i)dx_i = \frac{1}{n}\mu = \frac{1}{n}(np) = p$$

وعندما يكون عدد التجارب العشوائية كبيراً $(n \to \infty)$ يتقارب التوزيع الثنائي عادةً من التوزيع الطبيعي (نظرية) وبالتالي نجد أنّ المتغير الطبيعي المعياري (أو الدرجة المعيارية):

$$z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}} = \frac{\frac{x}{n}-p}{\sqrt{\frac{pq}{n}}} = \frac{\hat{p}-p}{\sqrt{pq/n}} \qquad (5.40)$$

حيث $q = 1-p$ و $\sigma^2 = npq$ للتوزيع الثنائي.

وعليه نجد أنّ z تقع ضمن مجال الثقة أو التقدير المجالي الطبيعي وبدرجة ثقة $p = (1-\alpha)\,100\%$ التالية: $-z_{\alpha/2} < Z < z_{\alpha/2}$ وبتبسيط المتباينة أعلاه نحصل على مجال الثقة أو التقدير المجالي للوسط p عندما يكون حجم العينة كبيراً.

$$\frac{x}{n} - z_{\alpha/2}\sqrt{\frac{pq}{n}} < p < \frac{x}{n} + z_{\alpha/2}\sqrt{\frac{pq}{n}} \qquad (5.41)$$

وعادةً ما يجري التغاضي عن مقدار الخطأ الناتج عن استبدال $\hat{p}$ ، $\hat{q}$ (التكرار النسبي لوقوع الحدث في العينة وعدم وقوعه) بدلاً من p ، q (التكرار النسبي لوقوع الحدث في المجتمع الإحصائي وعدم وقوعه) في طرفي المتراجحة أعلاه حيث نحصل على التقدير المجالي أو متراجحة الثقة المعدّلة التالية:

$$\frac{x}{n} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \frac{x}{n} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad (5.42)$$

وبنفس درجة الثقة السابقة $p = (1-\alpha)\ 100\%$ وحيث يمثل المقـدار $\varepsilon = Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$ أعلـى خطأ في التقدير. ومن الجدير بالإشارة إمكانية استخدام الخطأ في التقدير أعلاه لحساب حجـم العينة حيث نجد أنّ:

$$n = \frac{z^2_{\alpha/2}\ \hat{p}\hat{q}}{\varepsilon^2} \qquad (5.43)$$

وفي كافة الأحوال ونظراً للمرونة في تقدير $\hat{p}$ لحساب حجم العينة نجد أنّ: $0 < \hat{p} < 1$ لـذا فإنّ: $\hat{p}\hat{q} = \hat{p}(1-\hat{p}) = -(\hat{p}^2 - \hat{p}) = \frac{1}{4} - (\hat{p}^2 - \hat{p} + \frac{1}{4}) = \frac{1}{4} - (\hat{p} - \frac{1}{2})^2 \le \frac{1}{4}$ وعليه لو افترضنا بأنّ $\hat{p} = \frac{1}{2}$ (أو أنّ $\hat{q} = \frac{1}{2}$) فإننا نحصل على حجم عينة كبير جداً وبنفس درجة الثقة (أو الاحتمالية) كما في نص النظرية التالية:

نظرية ❷ (تحديد حجم العينة):

إذا كان $\hat{p}$ مقدّر للوسيط p فإنّ مقدار الخطأ في التقدير لا يزيد عن المقدار ε وبدرجـة ثقـة $p = (1-\alpha)\ 100\%$ إذا كان حجم العينة هو:

$$n = \frac{z^2_{\alpha/2}}{4\varepsilon^2} \qquad (5.44)$$

وهي نفس العلاقة (5.43) عند افتراض أنّ $\hat{p} = \hat{q} = \frac{1}{2}$.

مثال (19): أخذت عينة من 500 بطارية جافة ووجد بأنّ 160 منها ضعيفة أو تكـاد تكـون تالفة فما هو كل من التقدير النقطي والتقدير المجالي للبطاريات الضعيفة وبدرجة أو مسـتوى ثقة هو 95%.

الحل: التقـدير النقطـي للبطاريـات الضـعيفة هـو $\hat{p} = \frac{160}{500} = 0.32$ ومنـه نجـد أنّ: $\hat{q} = 1 - 0.32 = 0.68$ لاحظ أنّ $\hat{q}$ يمثل التقدير النقطي للبطاريات الجيدة غير الضعيفة.

لدينا $\alpha = 1 - 0.95 = 0.05$ ومن الجـداول نجـد أنّ: $z_{\alpha/2} < z_{0.025} = 1.96$ وبتعـويض المعلومات في متراجحة الثقة الطبيعية للاحتمال (5.42) نحصل على التالي:

$$0.32-1.96\sqrt{\frac{0.32\times0.68}{500}}<p<0.32+1.96\sqrt{\frac{0.32\times0.68}{500}}$$

$$0.28<p<0.36$$

مثال (20): حدد حجم العينة التي يجب أن تؤخذ في المثال أعلاه بحيث يمكننا تقدير قيمة المعلمة p (البطاريات الضعيفة) ضمن خطأ مقداره 0.02 ومستوى ثقة يساوي 95٪.

الحل: لنعتبر مرحلياً بأنّ حجم العينة هو 500 لذا فإنّ $\hat{p}=0.32$ كما في المثال السابق وعليه نجد أنّ حجم العينة الواجب أخذه هو: $n=\frac{(1.96)^2(0.32)(0.68)}{(0.02)^2}=2090$ وعلى هذا الأساس يمكننا القول بأنه إذا أخذنا حجم العينة 2090 فإننا سنكون على ثقة بمقدار 95٪ بأنّ المقدّر $\hat{p}$ سوف يختلف عن الوسيط الحقيقي p بمقدار لا يزيد عن 0.02 أي أنّ الحد الأعلى للتقدير هو $U=0.34$ وأنّ أدنى حد للتقدير هو $L=0.30$.

مثال (21): كم يجب أن يكون حجم العينة في المثال أعلاه على افتراض أن درجة الثقة بالخطأ 0.02 هي على الأقل $p=95\%$.

الحل:

$$n=\frac{z^2_{\alpha/2}}{4\varepsilon^2}=\frac{(1.96)^2}{4(0.02)^2}=2401$$

لاحظ الفرق بين حجمي العينة في كل من الحالتين.

5-8 تقدير وسيط الفرق بين نسبتين P_D لمجتمعين إحصائيين (proportion difference):

يمكن تقدير وسيط الفرق $P_D=P_1-P_2$ بين احتمالي وقوع حدث معين مثل e ضمن مجتمعين إحصائيين مختلفين للمقارنة بينهما تقديراً نقطياً أو تقديراً مجالياً عن طريق معرفة التوزيع التكراري للمقدّر $\hat{P}_D=\hat{P}_1-\hat{P}_2$

فإذا كان حجم العينتين n_1 ، n_2 المأخوذتين من المجتمعين المختلفين على الترتيب كبيراً بما فيه الكفاية نجد أنّ التوزيع الثنائي التكراري لكل من الاحصائيتين $\hat{P}_2$ ، $\hat{P}_1$ يأخذ شكل

المنحني الطبيعي الذي يتقارب مع التوزيع الثنائي عندئذٍ حيث المتوسط الحسابي للعينتين هـو $\mu_1 = P_1$ ، $\mu_2 = P_2$ والتباين هو $\sigma_1^2 = \frac{p_1 q_1}{n_1}$ ، $\sigma_2^2 = \frac{p_2 q_2}{n_2}$ على الترتيب حيث العينتـان عشوائيتان وغير مترابطتين مع بعضهما، وكما نجد أيضاً أنّ التوزيع التكراري الثنائي للمقدّر $\hat{P}_D = \hat{P}_1 - \hat{P}_2$ يتقارب من التوزيع الطبيعي وبمتوسط مقداره (التقدير النقطي).

$$\mu_D = \mu_{\hat{p}_1 - \hat{p}_2} = \hat{p}_1 - \hat{p}_2 \qquad (5.45)$$

وبتباين مقداره (التقدير النقطي):

$$\sigma_D^2 = \sigma_{\hat{p}_1 - \hat{p}_2}^2 = \frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2} \qquad (5.46)$$

وعلى أساس هذا التوزيع الطبيعي المتقارب مع التوزيع الثنائي (أو البرنولي) يمكننـا التأكيـد وبدرجة ثقة تساوي $P = (1-\alpha)\ 100\%$ على أنّ المتغير المعيـاري، (الدرجـة المعياريـة للفرق):

$$Z = \frac{\hat{P}_D - P_D}{\sigma_D} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \qquad (5.47)$$

يقع ضمن فترة الثقة أو التقدير المجالي التالي:

$$\hat{P}_D - Z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < P_D = P_1 - P_2 < \hat{P}_D + Z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \qquad (5.48)$$

حيث تمت الاستعاضة عن p_1،q_1 ، p_2،q_2 بتقـديراتها النقطيـة $\hat{p}_1$،$\hat{q}_1$ ، $\hat{p}_2$،$\hat{q}_2$ علـى الترتيب مع قبول بعض الخطأ الضئيل لهذه الاستعاضة.

وعلى نفس المنوال، عندما يكون حجمي العينتين (العشوائيتين) المأخوذتين مـن المجتمعـين الإحصائيين المختلفين صغيراً وليس كبيراً كما في الحالة السـابقة يمكننـا اسـتخدام توزيـع ستيودنت -t أي أنّ المتغير العشوائي يخضع لتوزيع t عندئذٍ. وبعبارة أخرى يمكننا اسـتخدام مجال الثقة التائي بدلاً من مجال الثقة الطبيعي السابق كالتالي:

$$\hat{P}_D - t_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < P_D < \hat{P}_D + t_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \qquad (5.49)$$

مثال (22): لتحديد الفرق بين نسبتي البطاريات الضعيفة لشركتين مختلفتين أخـذت عينتـان عشوائيتان من البطاريات بحجمين $n_1 = 300$ و $n_2 = 400$ ووجد أنّ في الأولى 64 بطارية

ضعيفة ووجد في الثانية 51 بطارية ضعيفة أحسب كلاً من التقدير النقطي والمجـالي لنسـب البطاريات الضعيفة وللفرق بين نسبتي الضعف لبطاريات كل من الشركتين وبدرجة ثقة 95٪.

الحل: التقدير النقطي $\hat{P}_1 = \frac{64}{300}$ ، $\hat{P}_2 = \frac{51}{400}$ ، $\hat{P}_D = \hat{P}_1 - \hat{P}_2 = \frac{256-153}{1200}$

ومنه: $\hat{q}_1 = \frac{236}{300}$ $\hat{q}_2 = \frac{349}{400}$ وبما أنّ درجـة الثقـة $P = 95\% = (1-\alpha)\,100\% \Leftarrow$

$\alpha = 0.05 \Leftarrow \frac{\alpha}{2} = 0.025$ ومـن جـداول التوزيـع المعيـاري نجـد أنّ: $z_{\alpha/2} = 1.96$

وبالتعويض في المتراجحة (5-48) نجد أنّ:

$$\frac{103}{1200} - 1.96\sqrt{50.35+44.50} < P_1 - P_2 < \frac{103}{1200} + 1.96\sqrt{50.35+44.50}$$

$$-19.003 < P_1 - P_2 < 19.175$$

ملاحظة ⑧: حجم العينة كبير أو صغير يعتمد على نوعية الدقة والثقة المطلوبة في المسـائل ففي المجتمعات المتماثلة التوزيع (الطبيعية) نسبياً نجد أنّ $n \geq 10$ قد تكون كافيـة ومقبولـة ولكن في كافة الأحوال أي حتى في عدم تماثل توزيع المجتمع الأم نجد أنّ $(n \geq 30)$ كافيـاً ومقبولاً. أمّا بالنسبة إلى t و χ^2 فهما يستخدمان بصورة مقبولة وكافية في كافة الأحوال حتى وإن لم يكن توزيع المجتمع الأم متماثلاً أو طبيعياً.

مخطط (5-1)

يوضح العلاقة بين خصائص المجتمعات الاحصائية الأم وحجم العينة n ومتوسطها $\overline{x}$ وتباينها S^2

	حالة التباين σ^2	نوعية التوزيع	حجم العينة	حدود المجتمع	نوعية الاختبار والدرجة المعيارية	الملاحظات
مسائل المتوسطات $\overline{x}$ بالنسبة لمتوسط المجتمع μ	تباين المجتمع الأم σ^2 معلوم أو مفترض	توزيع المجتمع الأم غير طبيعي	صغير $(n < 30)$	------	التجربة مرفوضة	------
			كبير $(n \geq 30)$	محدود وبدون إرجاع	$z = \frac{\overline{x}-\mu}{(\sigma/\sqrt{n})\sqrt{\frac{N-n}{N-1}}}$	الدرجة المعيارية بعد التصحيح
			كبير $(n \geq 30)$	لانهائي أو مع الإرجاع	$z = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$	توزيع $\overline{x}$ طبيعي
		توزيع المجتمع الأم طبيعي $N(\mu,\sigma^2)$	$(n < 30)$ أو $(n \geq 30)$	------	$z = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$	توزيع $\overline{x}$ طبيعي
	تباين المجتمع الأم σ^2 غير معلوم	توزيع المجتمع الأم غير طبيعي	صغير $(n < 30)$	------	التجربة مرفوضة	------
			كبير $(n \geq 30)$	------	$t = \frac{\overline{x}-\mu}{s/\sqrt{n}}$	استخدام توزيع ستيودنت
		توزيع المجتمع الأم طبيعي	صغير $(n < 30)$	------	$t = \frac{\overline{x}-\mu}{s/\sqrt{n}}$	استخدام توزيع ستيودنت
			كبير $(n \geq 30)$	------	$z = \frac{\overline{x}-\mu}{s/\sqrt{n}}$	استخدام التوزيع المعياري
مسائل التباين S^2 بالنسبة لتباين المجتمع σ^2		توزيع المجتمع الأم غير طبيعي	صغير $(n < 30)$	------	التجربة مرفوضة	------
			كبير $(n \geq 30)$	------	$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$	استخدام توزيع مربع كاي
		توزيع المجتمع الأم طبيعي	$(n < 30)$ أو $(n \geq 30)$	------	$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$	

5-9 تمارين عامة:

1- أ- أذكر شروط الأفضلية بين المقدّرين $\hat{\theta}_1$ ، $\hat{\theta}_2$ لتقدير الوسيط θ.

ب- أحسب الكفاءة النسبية للمقدّرين السابقين إذا علمت أنّ قيم كل منها كالتالي:

$\hat{\theta}_1$:	34	45	50	40	55
$\hat{\theta}_2$:	30	32	35	38	40

2- أخذت العينة {1, 5, 9, 8, 4, 0, 2, 4, 3} من مجتمع معين طبيعي التوزيع أحسب كلاً من التقدير النقطي والتقدير المجالي لمتوسط المجتمع المذكور (μ) وانحرافه المعياري (σ) وبدرجة ثقة 99٪ و 95٪.

3- إذا كان متوسط العمر الزمني لعينة حجمها $n = 30$ من المصابيح الكهربائية المصنعة بإحدى الشركات هو $\bar{x} = 780$ ساعة. وإذا علمنا بأنّ المصابيح المصنعة في هذه الشركة والممثلة للمجتمع الإحصائي تتوزع أعمارها الزمنية بصورة طبيعية تقريباً:

أ- أوجد متباينة الثقة لمتوسط الأعمار الزمنية لكافة المصابيح المصنوعة بتلك الشركة وبدرجة ثقة 96٪ إذا علمت أنّ الانحراف المعياري لأعمار تلك المصابيح هو $\sigma = 40$ ساعة.

ب- حدّد حجم العينة المطلوبة في التمرين أعلاه إذا كان احتمال الثقة بأنّ أعلى خطأ في متوسط العينة لا يزيد عن عشر ساعات وبدرجة ثقة 96٪.

جـ- أفترض عدم معرفة تباين مجتمع المصابيح في الشركة المذكورة وأحسب مجال الثقة (التقدير المجالي) على أساس ذلك.

4- عينة عشوائية ذات حجم $n = 20$ مختارة من مجتمع طبيعي، متوسطها $\bar{x} = 32.8$ وانحرافها المعياري $S = 4.51$ والمطلوب:

أ- أنشئ 95٪ مجال ثقة لـ μ

ب- ما هو حجم العينة المطلوب إذا أردنا أن نكون على ثقة 98٪ أن تقديرنا لـ μ لا يختلف بأكثر من 0.05.

5- عينة عشوائية مؤلفة من $n = 500$ سيدة، وجد أنّ $x = 160$ منهم يفضلون نوعاً معيناً من القماش، أوجد 95٪ مجال ثقة للنسبة الحقيقية للراغبات في هذا القماش بين مجتمع السيدات.

6- أ- أذكر العلاقة بين متوسط مربعات الخطأ والتباين مع التحليل.

ب- أُخذت عينة من تسع مفردات لفحص صلاحية مواد غذائية معلبة من شحنة مستوردة تحتوي على 950 علبة فوجد أنّ مدة صلاحية مفردات العينة كالتالي:

t_i : 30 , 28 , 24 , 26 , 29 , 20 , 27 , 25 , 31 فأحسب كلاً من التقدير المجالي لمتوسط صلاحية الشحنة المستوردة، والتقدير المجالي لتباين مدة صلاحية الشحنة المستوردة وبدرجة ثقة 95٪.

7- شركة لصناعة التلفاز تود معرفة نسبة المستهلكين الذين يستخدمون التلفاز الملون ومن عينة عشوائية ذات مائة مستهلك وجد أن 60٪ منهم يملكون أجهزة ملونة أحسب مجال الثقة وبدرجة 95٪ لنسبة المجتمع الإحصائي P لمن يمتلكون أجهزة ملونة.

8- تُستوقف آلة لصناعة كرات الحركة الحديدية (ball bearings) كل مرة لفحص تباين أقطارها وسلامة الآلة، فإذا أخذت عينة بحجم $n - 31$ ووجد أنّ تباين أقطارها $S^2 - 0.94$ ملم: أ- كوّن مجال الثقة (التقدير) وبدرجة 95٪ لتباين مجتمع الكرات لهذه الآلة وعلى افتراض أنّ التوزيع طبيعياً.

ب- على افتراض أنّ الآلة تعمل بصورة طبيعية (سليمة) وأنّ تباين قطر الكرات المنتجة هو $\sigma^2 = 0.50$ ملم فهل نستنتج من العينة أعلاه بأنّ الآلة غير سليمة.

9- إذا كان 5٪ من 10000 منتج لأحدى الشركات الصناعية معطوباً وأراد جهاز الرقابة التأكد من صحة نسبة العطل، فإذا كانت معلومات الرقابة الأولية تفيد بأنّ عينة من 288 منتجاً تفي بالغرض بدرجة ثقة 95٪ ومقدار الخطأ $\varepsilon = \pm 0.25$ فإذا تم تغيير مقدار الخطأ المسموح به إلى $\varepsilon = \pm 0.05$ ودرجة الثقة نفسها فما هو حجم العينة الواجب خذها من المنتجات.

10- تريد إحدى شركات التاكسي شراء عجلات جديدة نوع A أو نوع B لأسطولها من السيارات ولتقدير الفرق بين النوعين أخذت عينتان من 12 عنصراً من كل نوع فإذا وجد أنّ استخدام العجلات حتى التلف يعطي البيانات التالية:

نوع A : $\bar{x}_1 = 36300km$ $S_1 = 5000km$

نوع B : $\bar{x}_2 = 38100km$ $S_2 = 6100km$

فأوجد مجال الثقة أو التقدير للفرق $\mu_D = \mu_1 - \mu_2$ للمجتمع الإحصائي وبدرجة 95٪ معتبراً أنّ التوزيع من النوع الطبيعي.

11- أ- عرّف مفهوم عدم الانحياز للمقدّر $\hat{\theta}$.

ب- وضّح بأنّ كلاً من الإحصائيتين، التباين التجميعي وتباين العينة:

$$S^2 = \frac{\sum(x_i - x)^2}{n-1} \text{ ، } S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

هما مقديرين غير منحازين للتباين σ^2.

12- إذا كانت:

$$f_X(x) = \begin{cases} 2(1-x) & 0 < x < 1 \text{ إذا كانت} \\ 0 & \text{لغير ذلك} \end{cases}$$

فأحسب التقدير النقطي المتوقع لكل من μ ، σ^2.

13- أخذت عينتان ذواتا حجم $n_1 = 9$ ، $n_2 = 16$ من مجتمعين إحصائيين طبيعيين وغير مترابطين بحيث كان $\bar{x}_1 = 64$, $\bar{x}_2 = 59$ و $S_1 = 6$, $S_2 = 5$ أوجد مجال الثقة (التقدير) للفرق بين المتوسطين $\mu_1 - \mu_2$ وبدرجة احتمال 95٪ وعلى افتراض أنّ $\sigma_1 = \sigma_2$.

14- أخذت عينة عشوائية ذات حجم $n_1 = 25$ من مجتمع طبيعي التوزيع وبانحراف معياري $\sigma_1 = 5$ ووجد أنّ متوسطها $\bar{x}_1 = 80$. كما أُخذت عينة عشوائية ثانية حجمها $n_2 = 36$ من مجتمع آخر طبيعي التوزيع وذو انحراف معياري $\sigma_2 = 3$ ووجد أنّ متوسطها هو $\bar{x}_2 = 75$ فأحسب مجال الثقة (التقدير) لفرق الوسيطين $\mu_1 - \mu_2$ وبدرجة ثقة 94٪.

15- وضّح بأنّ إحصائية التقدير التجميعية S^2 للوسيط σ^2 هي مقدّر غير منحاز:

حيث :
$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

16- الإحصائيات التالية توضح قوة التحمل لنوعين من مواد التغليف (قدم-باوند) أوجد مجال الثقة للفرق بين المتوسطين وبدرجة 98٪.

A:	1.25	1.16	1.33	1.15	1.23	1.20	1.32	1.28	1.2
B:	0.89	1.01	0.97	0.95	0.94	1.02	0.98	1.06	0.98

17- إذا كان متوسط عمر بطاريات السيارات المصنعة في أحد المصانع هو 3 سنوات وتباينها هو سنة واحدة. أخذت عينة من خمس بطاريات ووجد أنّ عمرها الزمني بالسنوات هو على الترتيب 1.9، 2.4، 3.0، 3.5، 4.2.

والمطلوب: أوجد مجال الثقة للوسيط σ^2 وبدرجة احتمال 95٪ وتحقق ما إذا كان تباين السنة الواحدة صحيحاً ومقبولاً أم لا على افتراض أن التوزيع التكراري لمجتمع البطاريات في ذلك المصنع من النوع الطبيعي.

18- أخذت عينة من 15 عنصراً ووجد أنّ S=3 فأحسب مجال الثقة أو التقدير للانحراف المعياري σ للمجتمع الإحصائي وبدرجة ثقة 90٪ وعلى افتراض أنّ توزيع المجتمع هو توزيع طبيعي.

الفصل السادس

اختبار الفرضيات

6-1 – تمهيد.

6-2 – اختبارات حول المتوسط (μ).

6-3 – اختبارات حول التباين $\left(\sigma^2\right)$.

6-4 – اختبارات النسب والاحتمال.

6-5 – اختبار التوافق والتجانس.

6-6 – اختبار الاستقلالية والترابط.

6-7 – اختبار الإشارة.

6-8 – تمارين عامة.

اختبار الفرضيات
(Test Hypothesis)

6-1 تمهيد:

كثيراً ما نواجه الصعوبات في اتخاذ القرار المناسب في عملية قبـول أو رفـض مسـألة أو فرضية معينة أو حتى تقدير معين، فالمهندس الذي يتوجب عليه أن يتخذ قراراً بأفضلية نوعين من أنواع المحركات أو أفضلية أو تكافؤ نوعين من أنواع مبدلات الحركة بناءً على بيانـات ومشاهدات مأخوذة عن عينة عشوائية لا بد وأن يكون على درجة عالية من الثقة ليقرر ما إذا كان النوع الأفضل من المحركات ضمن العينة يمثل النوع الأفضل مـن المحركـات ضـمن المجتمع الإحصائي بالكامل أم لا.

تعريف (1):

الفرضية الإحصائية والرفض والقبول: تعرف الفرضية الإحصائية على أنها عبارة أو تصور يتعلق بكامل المجتمع الإحصائي الأم مأخوذة عن عينة عشوائية منه. أمّا الـرفض والقبـول للفرضية الإحصائية والصح والخطأ فهي تعابير تستخدم في هذا المجال. ولا بد من الإشـارة هنا إلى أنّ رفض الفرضية لا يعني خطأها أمّا قبول الفرضية الإحصائية فلا يعني بالضرورة صحتها ولكنه يعني عدم وجود الحجة والدليل القاطع لرفضها. تصنف الفرضيات الإحصـائية إلى نوعين هما:

1- **فرضية العدم (Null hypothesis):** والتي يرمز لها بالرمز H_0 عادةً وتعني انعدام الاختلاف والتباين أي التساوي (=) كما تسمى أحياناً بفرضـية الـرفض أي رفـض الاختلاف.

2- فرضية البديل (Alternative hypothesis): والتي يرمز لها بالرمز H_1 عادة وتعني وجود الاختلاف والتباين وعدم انعدامهما، كما تسمى أحياناً بفرضية القبول أي القبول بوجود الاختلاف والتباين (<،>) (≠).

ويلاحظ من التصنيف السابق بأنّ القبول بالعدم يعني رفض الافتراض البديل وبالعكس نلاحظ بأنّ قبول البديل يعني رفض التساوي وقبول عدم التساوي. إنّ عدم التساوي في فرضية البديل يعني واحداً من حالات ثلاث هي:

أ – قبول التباين الأكبر (<) وتسمى الفرضية عندئذ بذات الطرف الواحد.

ب– قبول التباين الأصغر (>) وتسمى الفرضية عندئذ بذات الطرف الواحد.

جـ – قبول كلا التباينين أعلاه (≠) وتسمى الفرضية عندئذ بذات الطرفين.

6-1-1 كيفية صياغة الفرضية الإحصائية: تُصاغ الفرضية الإحصائية وفقاً للمعلومات أو البيانات المأخوذة من العينة العشوائية عن المجتمع الإحصائي الأم أو المجتمعات الإحصائية قيد الدراسة وعادةً ما تُصاغ فرضية العدم على ضوء المفهوم المراد رفضه بحيث تكون محددة وواضحة ووحيدة غير متعددة بينما تُصاغ فرضية القبول عادةً على ضوء المفهوم المراد قبوله وقد تكون متعددة المضمون. فمثلاً لو أردنا اختبار نوع جديد من المحركات (A) المصنوعة في شركة معينة واستخدامه في الأسواق ومقارنته بالنوع القديم (B) من المحركات المصنوعة في نفس الشركة أو شركة أخرى والذي تتراوح نسبة العطل فيه بمقدار $p_1=25\%$ خلال ثلاثة أشهر. ولغرض اتخاذ القرار في استخدام المحرك الجديد من عدمه يتوجب أن تكون نسبة العطل (p) فيه أصغر من$p_1=25\%$ والتي تمثل نسبة العطل في المحركات (B) وبمقدار معقول. وعليه لو أخذنا عينة من 60 محرك قديم B فإنّ نسبة العطل المتوقعة هي 15 محرك خلال ثلاثة أشهر ولو أخذنا نفس العدد من المحركات الجديدة A فإنّ نسبة العطل فيها يجب أن يكون أقل من 15 محركاً خلال ثلاثة أشهر ولنقل مثلاً 13 محركاً. وعلى الرغم من تقارب النسبة فإنّ التساؤل هو حول ترك أو رفض المحركات القديمة وقبول المحركات الجديدة بدلاً عنها. وعلى مضمون ما تقدم تصاغ فرضية العدم على أساس تساوي نسب العطل في كل من A، B أي أنّ:

$$H_0 : P = p_1 = 25\%$$

وهي النسبة التي نرغب في تقليصها. أمّا فرضية القبول بالاختلاف (أو الفرضية البديلة) فتصاغ بالشكل التالي: $H_1 : P< p_1 = 25\ \%$ وبديهي أن تعتبر النسبة $p_1=25\%$ حداً فاصلاً أو حرجاً (critical value) بين الرفض والقبول أي أنّ العدد $x_0=15$ هو الحد الحرج الفاصل بين مجموعتي الأعداد المقبولة والمرفوضة. فلو افترضنا أن X يرمز إلى عدد المحركات التي تتعطل خلال الفترة الزمنية المحددة فإنّ X يمثل متغيراً عشوائياً يأخذ القيم التالية: $0\leq x\leq 60$ ومنه نجد أنّ مجموعة الرفض هي: $R_1=\{x_i:15\leq x_i\leq 60\}$ بينما نجد أن مجموعة القبول هي: $R_2=\{x_i:0\leq x_i<15\}$. وبعبارة أخرى يتم رفض H_1 (أي قبول H_0) إذا كان المتغير $X>x_0=15$ بينما نرفض H_0 (أي نقبل H_1) إذا كان المتغير X (الذي يمثل عدد المحركات التي تتعطل خلال الفترة المحددة) أصغر من $x_0=15$ محركاً. ومن الجدير بالذكر هنا أنّ الحد الحرج الفاصل $x_0=15$ يمكن تغييره إلى أقل من ذلك وفقاً لشروط العمل والجودة مثل $x_0=14$ أو $x_0=13$ مثلاً.

6-1-2 أخطاء الفرضيات وحسابات الخطأ: في خضم قرارات القبول والرفض نلاحظ بأنّ هنالك نوعين من الأخطاء هما:

1- الخطأ من النوع الأول والذي يتم فيه رفض فرضية العدم H_0 وهي في الواقع صحيحة. ويرمز لهذا الخطأ عادةً بالرمز α.

2- الخطأ من النوع الثاني والذي يتم فيه قبول فرضية العدم H_0 وهي في الواقع خاطئة ويرمز لهذا الخطأ عادة بالرمز β والجدول التالي يوضح الحالات الأربع لاتخاذ القرار:

جدول (6-1)

حالة H_0 / القرار المتخذ	H_0 صحيحة	H_0 خاطئة
قبول H_0	قرار صحيح مستوى الثقة $p=(1-\alpha)$	قرار خاطئ من النوع الثاني (β)
رفض H_0	قرار خاطئ من النوع الأول (α) **α=level of significance**	قرار صحيح قوة الاختبار = $(1-\beta)$

ففي مثال المحركات نجد أنّ الخطأ من النوع الأول هو رفض H_0 وقبول المحركات الجديدة على الرغم من أنّ p=25٪ وبالعكس نجد أنّ الخطأ الثاني هو قبول H_0 ورفض المحركات الجديدة على الرغم من أن نسبة العطل فيها هو أقل من 15 محركاً (p<25٪). وبديهي أنّ الوقوع في مثل هذه الأخطاء قد لا يكون على درجة كبيرة من الأهمية ولكنه محتمل.

ولتجنب الوقوع في الخطأ من النوع الثاني $\beta \to 0$ أي قبول H_0 وهي خاطئة يتوجب زيادة α أي زيادة منطقة أو احتمال الرفض وهذه الزيادة تعني زيادة نسبة الوقوع في الخطأ الأول وبالعكس. ومن جانب آخر يمكننا اختزال الوقوع في الخطأين الأول والثاني كلما ازداد حجم العينة العشوائية حيث يتقارب عندئذٍ التوزيع الثنائي من التوزيع الطبيعي.

مثال (1): تدعي إحدى شركات التصنيع بأنّ متوسط عمر بطارياتها الجافة والقابلة للشحن هو 68 ساعة متواصلة وعلى أساس أنّ حجم العينة المأخوذة هو n=36 بينما الانحراف المعياري المفترض للمجتمع الإحصائي هو σ=3.6. كوّن الفرضيات المناسبة وأحسب احتمال الوقوع في الخطأين الأول α والثاني β.

الحل: نلاحظ بأنّ حجم العينة المأخوذة n>30 لذا فإنّ توزيع متوسط العينات يتقارب مع التوزيع الطبيعي $N(\mu, \sigma^2)$ وأنّ: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.6}{\sqrt{36}} = 0.6$ كما نلاحظ ادعاء الشركة وعلى أساسه نتساءل عن صحة أو عدم صحة الادعاء وعليه تكون فرضيتي العدم والقبول كالتالي:

1- فرضية اللا اختلاف مع متوسط العينة $H_0 : \mu = 68 = \bar{x}$

2- الاختلاف مع متوسط العينة $H_1 : \mu \neq 68$

والفرضية الأخيرة من ذات الطرفين $\mu > 68$ و $\mu < 68$

ولما كان التوزيع طبيعياً فإنّ أكثر العناصر تتمركز حول المتوسط الحسابي ووفقاً للمتراجحة التالية:

$$67.4 = 68 - 0.6 < \bar{x} < 68 + 0.6 = 68.6$$

إنّ قبول H_0 يعني وقوع متوسط المجتمع الإحصائي للبطاريات قريباً من 68 أو يساويه أمّا قبول H_1 فيعني وقوع المتوسط μ بعيداً وخارج الفترة السابقة أي أنّ $\mu > 68.6$

أو $67.4 > \mu$. أمّا احتمال الوقوع في الخطأ من النوع الأول α أي احتمال رفض H_0 وهي صحيحة أو احتمال وقوع $\bar{x}$ في منطقة الرفض وخارج الفترة السابقة فيحسب كالتالي:

$$\alpha = P(\bar{X} < 67.4 \ , \ \mu = 68) + P(\bar{X} > 68.6 \ , \ \mu = 68)$$

وبالتحويل إلى الصورة المعيارية الطبيعية نجد أنّ:

$$\alpha = P(Z < \frac{67.4-68}{0.6}) + P(Z > \frac{68.6 \ -68}{0.6}) = P(Z < -1) + P(Z > +1)$$
$$= 2P(Z < -1) = 2(0.1587) = 0.3674$$

وهو احتمال كبير نسبياً ويستدعي رفض H_0 على الرغم من أنها صحيحة ولتقليل هذا الاحتمال α أي تقليل الوقوع في الخطأ من النوع الأول سوف نزيد حجم العينة وهذا يعني زيادة منطقة القبول وتقليص منطقة الرفض. فلو افترضنا أنّ حجم العينة $n = 64$ مثلاً فإنّ الدرجات المعيارية والاحتمال α يأخذان الصورة التالية:

$$\alpha = P(Z < z_{\alpha/2}) + P(Z > z_{\alpha/2}) = P(Z < \frac{67.4-68}{0.45}) + P(Z > \frac{68.6-68}{0.45})$$
$$= P(Z - 1.33) + P(Z > 1.33) = 2P(Z < -1.33) = 0.1836$$

وهو احتمال أصغر من السابق ولكنه كبير نسبياً أيضاً ويستدعي رفض H_0 ، وهكذا نلاحظ ضرورة زيادة حجم العينة كي يتم تقليص هذا الخطأ α. والحقيقة إنّ تقليص α لا يكفي وحده لضمان جودة الاختبار بل يتوجب حساب β أيضاً ولعدد من الفرضيات البديلة H_1. فلو افترضنا في هذا المثال أنّ $\mu = 70$ مثلاً بدلاً من 68 فإنّ حساب احتمال الوقوع في الخطأ من النوع الثاني β أي قبول H_0 وهي خاطئة نجد أنّ:

$$\beta = P(67.4 \le \bar{X} \le 68.6, \mu = 70)$$
$$= P(\frac{67.4-70}{0.45} \le Z \le \frac{68.6-70}{0.45}) = P(-5.78 \le Z \le -3.11)$$
$$= P(Z \le -3.11) - P(Z \le -5.78)$$

ومن الجداول الخاصة نجد أنّ:

$$\beta = 0.0009 - 0.00 = 0.0009$$

وهو احتمال ضئيل جداً للوقوع في الخطأ من النوع الثاني.

وبنفس الأسلوب يمكننا حساب احتمالات الوقوع في الخطأ من النوعين الأول والثاني. ولو افترضنا بأنّ $\mu = 68.5$ لوجدنا أنّ احتمال الوقوع في الخطأ الثاني كبير جداً بعكس احتمال الوقوع في الخطأ من النوع الأول الذي سيكون ضئيلاً جداً. وهي مسألة متروكة للطالب.

مثال (2): صرحت إحدى شركات تصنيع السجائر بأنّ نسبة النيكوتين في منتجاتها لا تزيد عن 2.5 ملي غرام للسيجار الواحد. أكتب الفرضيات المناسبة لمنح الشركة رخصة الاستمرار في التصنيع.

الحل: لمنح الشركة رخصة الاستمرار في التصنيع يتوجب أن تكون نسبة النيكوتين في المجتمع الإحصائي أقل أو مساوية للنسبة $\bar{x} = 2.5$ التي تمثل متوسط العينة أمّا الفرضية البديلة فبخلاف ذلك: أي أنّ:

فرضية اللا اختلاف (العدم) $H_0 : \mu = 2.5 = \bar{x}$

الفرضية البديلة $H_1 : \mu > 2.5$

والاختبار ذو طرف واحد لأنّ $\mu \leq 2.5$ لا يسبب مشكلة في منح الرخصة لأنّ H_0 يمثل الحد الأقصى لمنح الرخصة.

مثال (3): تدعي إحدى الشركات بأنّ 97٪ من سياراتها عالية الكفاءة تسير 10 كيلومترات لكل لتر من البنزين أكتب الفرضيات المناسبة لتحقيق صحة الادعاء.

الحل: للتحقيق في صحة ادعاء الشركة أي صحة النسبة المعطاة من قبل الشركة من عدمها يتوجب أن تكون فرضيتي الاختبار كالتالي:

$H_0 : P = 97\%$ $H_1 : P \neq 97\%$

والاختبار في هذه الحالة ذو طرفين لأنّ رفض ادعاء الشركة أي رفض H_0 ينتج من حالتين هما $P > 97\%$ وكذلك $P < 97\%$ وليس فقط عن احداهما.

3-1-6 الخطوات الستة للاختبار:

يمكننا صياغة أسلوب اختبار الفرضيات بالخطوات الستة التالية:

1- تحديد فرضية العدم H_0 واختيار الفرضية البديلة المناسبة H_1 من بين الحالات الثلاث: طرف واحد $\theta > \theta_0$ أو $\theta < \theta_0$ ، أو ذات الطرفين $\theta \neq \theta_0$.

2- تحديد احتمالي الرفض α والقبول $p = (1-\alpha)\ 100\%$

3- تحديد نوعية التوزيع واختيار الإحصائية المناسبة (Z، t، F، χ^2...) وإيجاد الدرجة (أو القيمة) الحرجة لها من الجداول الخاصة بها.

4- حساب قيمة الإحصائية θ_c المحددة في الفقرة (3) من بيانات العينة حيث:

$$\theta_c = \{z_c, t_c, f_c, \chi^2_c, ...\}$$

5- التحليل ومقارنة قيمة الإحصائية المحسوبة θ_c من العينة بالدرجة الحرجة (فقرة 3).

6- القرار: رفض H_0 وقبول H_1 إذا كانت قيمة الإحصائية المحسوبة θ_c من العينة تقع ضمن المنطقة الحرجة وخارج منطقة القبول. أو قبول H_0 ورفض H_1 بخلاف ذلك أي عندما تقع قيمة الإحصائية θ_c المحسوبة من العينة داخل منطقة القبول.

6-2 اختبارات حول المتوسط (μ):

تتلخص مسألة اختبار المتوسط μ لمجتمع إحصائي معلوم التباين σ^2 على أنها تساوي قيمة محددة μ_0 أم لا بصياغة الفرضيات الإحصائية كالتالي:

$$H_0 : \mu = \mu_0 \qquad\qquad H_1 : \mu \neq \mu_0$$

إنّ الإحصائية غير المنحازة والكفوءة الكفيلة بتحقيق هذا الاختبار هي متوسط العينة العشوائية $\overline{x}$ المأخوذة من المجتمع الإحصائي نفسه وذات الحجم n وكما نعرف بأنّ توزيع هذه الإحصائية يأخذ الصورة الطبيعية $N(\mu_{\overline{x}}, \sigma^2_{\overline{x}})$ تقريباً حيث $\mu_{\overline{x}} = \mu$ ، $\sigma^2_{\overline{x}} = \dfrac{\sigma^2}{n}$.

وباختيار أو تحديد مستوى ثقة معين مثل α (منطقة الرفض عادةً) يمكننا تحديد نقطتين حرجتين على طرفي التوزيع هما $\overline{x}_1$ ، $\overline{x}_2$ مثلاً بحيث أنّ الفترة

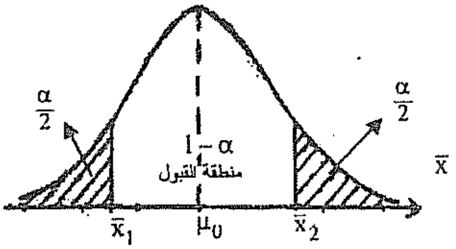
$$\overline{x}_1 < \overline{X} < \overline{x}_2 \qquad (6.1)$$

تمثل الفترة أو منطقة القبول $(1-\alpha)$ من التوزيع الطبيعي أمّا $\overline{X} > \overline{x}_2$ أو $\overline{X} < \overline{x}_1$ فتمثلان المنطقة الحرجة أو منطقتي الرفض اليمنى واليسرى من التوزيع الطبيعي كما في الشكل (6-1) ولمّا كان التوزيع طبيعياً، يمكننا تحويل فترة القبول السابقة إلى فترة قبول طبيعية معيارية على منحني التوزيع الطبيعي المعياري للدراسة وذلك بحساب الدرجة المعيارية الحرجة $\pm z_{\alpha/2}$ المناظرة للحدين $\overline{x}_1$ ، $\overline{x}_2$ على الترتيب كالتالي:

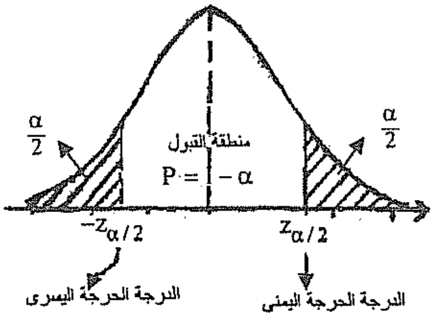
$-z_{\alpha/2} = \dfrac{\overline{x}_1 - \mu_0}{\sigma/\sqrt{n}}$ ، $z_{\alpha/2} = \dfrac{\overline{x}_2 - \mu_0}{\sigma/\sqrt{n}}$ وبالتالي نجد أنّ فترة القبـول الطبيعيـة المعياريـة المناظرة لفترة القبول السابقة هي:

$$-z_{\alpha/2} \le Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \le +z_{\alpha/2} \qquad \textbf{(6.2)}$$

كما في الشكل (6-2).



الشكل (6-1)



الشكل (6-2)

فإذا كانت $\overline{x}$ تقع ضمن فترة القبول (6.1) فإن z ستقع أيضاً ضمن فتـرة القبـول الطبيعيـة المعيارية (6.2) المناظرة للفترة الأولى وبالعكس وعلى هذا الأساس نستنتج صحة الفرضـية H_0 ويتم قبولها ورفض H_1 وبالمثل لو وقعت z فـي منطقـة الـرفض. أيّ $Z > z_{\alpha/2}$ أو $Z < -z_{\alpha/2}$ فإنّ $\overline{x}$ تقع كذلك وبالعكس وعلى هذا الأساس نستنتج عدم صحة فرضية العـدم $H_0 : \mu = \mu_0$ ونرفضها ونقبل عندئذ بالفرضية البديلة $H_1 : \mu \ne \mu_0$.

الاختبار السابق ذو الطرفين يناظر ويكافئ عملية حساب فترة الثقـة $100\% (1-\alpha)$ لوقـوع متوسط المجتمع الإحصائي μ ضمنها وقبول H_0 وبخلاف ذلك: أي وقوع μ خارج فترة الثقة ورفض H_0.

ملاحظة (1): يتم استبدال الدرجة المعيارية Z بالدرجة التائية t في حالة $n < 30$ كمـا يـتم استبدال التباين σ^2 بتباين العينة العشوائية S^2 عند عدم معلوميه التباين للمجتمع الأم والجداول التالية توضح كلاً من الفرضيات الإحصائية والإحصائيات المناظرة المستخدمة لاختبار تلـك الفرضيات.

جدول (6-2)

اختبارات المتوسط الحسابي μ للمجتمع الإحصائي معلوم التبـــاين σ^2 (أو مجهـــول التبـــاين و $n \geq 30$)

شروط ومجالات الرفض	إحصائية الاختبار	الفرضيات
$\lvert Z \rvert > z_{\alpha/2}$		$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
$Z < -z_\alpha$	$z_c = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$
$Z > z_\alpha$		$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
$\lvert Z \rvert > z_{\alpha/2}$		$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$
$Z < -z_\alpha$	$z_c = \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$
$Z > z_\alpha$	(مجتمعين إحصائيين)	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$

إذا كان تباين المجتمع σ^2 مجهولاً و $n \geq 30$ يبقى التوزيع Z هو المستخدم مع استخدام S^2 بدلاً من σ^2

جدول (3-6)

اختبارات المتوسط الحسابي μ للمجتمع الإحصائي غير معلوم التباين σ^2 و $n < 30$

الفرضيات	إحصائية الاختبار	شروط ومجالات الرفض
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$		$\lvert t \rvert > t_{\alpha/2, n-1}$
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$t_c = \dfrac{\overline{x} - \mu_0}{S/\sqrt{n}}$	$t < -t_{\alpha, n-1}$
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$		$t > t_{\alpha, n-1}$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$t_c = \dfrac{\overline{x}_1 - \overline{x}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ $\nu = n_1 + n_2 - 2$	$\lvert t \rvert > t_{\alpha/2, \nu}$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 < \mu_2$	$t_c = \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	$t > -t_{\alpha, \nu}$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 > \mu_2$	$\nu = \dfrac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1+1} + \dfrac{(S_2^2/n_2)^2}{n_2+1}} - 2$	$t > t_{\alpha, \nu}$

مثال (4): أخذت عينة من مائة مصباح عادي منتجة في شركة معينـــة فوجـد أنّ متوســط أعمارها هو 71.8 ساعة بانحراف معياري قدره S=8.9 ساعة فهل يعني ذلك بـأنّ متوسـط أعمار المصابيح المنتجة لدى هذه الشركة أطول من 70 ساعة وهل تمنح الشركة ترخيصـاً للاستمرار في الإنتاج أم لا؟ استخدم مستوى الأهمية (احتمال الرفض) $\alpha = 0.05$

الحل: تصاغ الفرضيات حسب المعطيات باستخدام الخطوات الستة كالتالي:

1- تحديد الفرضيات: $H_0:\mu=70$ ، $H_1:\mu>70$ الاختبار ذو طرف واحد لأن الفرض $\mu<70$ غير مقبول 2- لدينا $\alpha=0.05$ وبما أنّ $n=100>30$ لذا نستخدم التوزيع الطبيعي المعياري، وبما أنّ تباين المجتمع غير معروف لذا نستخدم $S=8.9$ بدلاً عن σ 3- تحديد الدرجة الحرجة ومنطقة الرفض: بما أنّ $\alpha=0.05$ لذا فإنّ الدرجة الطبيعية المعيارية الحرجة المناظرة لها هي $z_\alpha=1.645$ (من الجداول) وبالتالي فإنّ منطقة الرفض هي: $Z>1.645$ 4- حساب الدرجة المعيارية من العينة: وعليه فإنّ الدرجة المعيارية المقابلة لمعطيات العينة هي: $z_c=\dfrac{\overline{x}-\mu}{S/\sqrt{n}}=\dfrac{71.8-70}{8.9/10}=2.02$ 5- التحليل:

لما كانت القيمة $z_c=2.02$ أكبر من القيمة المعيارية الحرجة وبالتالي فهي تقع في منطقة الرفض أي: $z_c=2.02>z_{\alpha/2}=1.645$ 6- القرار: وعليه فإننا نستنتج رفض الفرضية H_0 أي رفض متوسط الأعمار $\mu=70$ وقبول الفرضية البديلة $H_1:\mu>70$ وعليه تمنح الشركة ترخيصاً للعمل والاستمرار في إنتاج هذا النوع من المصابيح.

مثال (5): يُعتقد بأنّ سرعة جريان سائل كثيف في وسط غير خشن هي 50 ملم/دقيقة ولكن عندما أخذت عينة عشوائية مكونة من 12 سائل وبنفس درجة الكثافة وجد أنّ سرعتها في نفس الوسط الخشن هو 42 ملم/دقيقة وبانحراف معياري هو 11.9 ملم/دقيقة. اختبر صحة الفرضية القائلة بأنّ متوسط سرعة الجريان هي أقل مما يعتقد سابقاً مستخدماً مستوى للأهمية أ- $\alpha=0.05$ ب- $\alpha=0.01$ ومعتبراً أنّ التوزيع طبيعي.

الحل: باستخدام الخطوات الستة نجد: 1- $H_0:\mu=50$ ، $H_1:\mu<50$ لاحظ هنا بأنه قد تم ذكر الفرضية البديلة في السؤال نفسه ولو كان النص غير ذلك مثلاً: اختبر صحة الاعتقاد السائد. فإنّ الفرضية البديلة ستكون $H_1:\mu\neq 50$ عندئذ. 2- لدينا $\alpha=0.05$ و $\alpha=0.01$ وبما أنّ حجم العينة $n=12<30$ والتوزيع الأم طبيعي فإننا نستخدم التوزيع -t والدرجة المعيارية التائية وكذلك S بدلاً من التباين σ لعدم معلوميته. 3- تحديد منطقة الرفض: المنطقة الحرجة التائية (منطقة الرفض) المناظرة لمستوى الأهمية $\alpha=0.05$ هي: $t<-1.796$ والمنطقة الحرجة التائية المناظرة لمستوى الأهمية $\alpha=0.01$ هي: $t<-2.718$ 4- وبما أنّ حجم العينة هي $n=12$ لذا فإنّ درجات الحرية $v=n-1=11$ وبالتالي فإنّ الدرجة المعيارية التائية للعينة:

5- التحليل: نلاحظ بأنّ الدرجة التائية المحسوبة $t_c = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{42-50}{11.9/\sqrt{12}} = -2.33$

$t_c = -2.33$ أصغر من -1.796 أو أنّ $t = -2.33 < t_{0.05,11} = -1.796$ لذا فهي تقع ضمن منطقة الرفض 6- القرار: وعليه نرفض الفرضية $H_0 : \mu = 50$ ونقبل الفرضية البديلة $H_1 : \mu < 50$ وبدرجة ثقة 95٪. كما نلاحظ أيضاً بأنّ t_c تقع ضمن منطقة القبول عندما تكون $\alpha = 0.01$ أي أنّ $t_c = -2.33 > t_{0.01,11} = -2.718$ وعليه نقبل الفرضية $H_0 : \mu = 50$ ونرفض الفرضية البديلة $H_1 : \mu < 50$ عندما تكون $\alpha = 0.01$ حيث اتخذ هذا القرار بدرجة ثقة99٪.

مثال (6): تدعي شركة لتصنيع أجهزة الهاتف المنزلي النقال بأنّ متوسط نصف قطر مدى هواتفها μ هو 15 متراً وبانحراف معياري σ قدره 0.8 متر اختبر مصداقية الادعاء بأخذ عينة عشوائية من 50 هاتفاً نقالاً للشركة مفترضاً أن متوسطها $\bar{x} = 14.8$ ومستخدماً درجة ثقة $1-\alpha = 99\%$

الحل: بإتباع نفس الخطوات السابقة السداسية المراحل نجد أنّ: 1- $H_0 : \mu = 15$ ، $H_1 : \mu \neq 15$ 2- النسبة الحرجة: لدينا $\alpha = 1-0.99 = 0.01$ 3- تحديد منطقة الرفض: بما أنّ $n > 30$ لذا فالتوزيع يتقارب من التوزيع الطبيعي المعياري والدرجة المعيارية الحرجة $\pm z_{\alpha/2}$ المناظرة إلى النسبة الحرجة α هي: $Z < -z_{\alpha/2} = -2.575$ و $Z > z_{\alpha/2} = 2.575$ حيث $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$. 4- حساب الدرجة المعيارية من العينة

5- التحليل وتحديد موقع الدرجة المعيارية: $z_c = \frac{14.8-15}{0.8/\sqrt{50}} = \frac{-0.2}{0.8/7.071} = -1.77$

بمقارنة الدرجة المعيارية z_c الناتجة من العينة العشوائية بمنطقة الرفض نلاحظ بأنّ $z_c = -1.77$ تقع خارج منطقتي الرفض أي داخل منطقة القبول لأنّ: $-2.575 < z_c = -1.77 < 2.575$ 6- اتخاذ القرار: بما أنّ z_c تقع داخل منطقة القبول لذا نقبل بادعاء الشركة H_0 ونرفض H_1 أي أنّ $\mu = 15$ فعلاً.

مثال (7): أخذت عينة من 12 طالباً يدرسون مادة الإحصاء بالطريقة المتعارف عليها بينما أخذت عينة أخرى من 10 طلاب يدرسون نفس المادة بالطريقة المتلفزة غير المباشرة وفي الاختبار النهائي وجد أنّ متوسط العينة الأولى هو 85 درجة وبانحراف معياري مقداره 4 درجات بينما كان متوسط العينة الثانية هو 81 درجة وبانحراف معياري هو 5 درجات. اختبر فرضية تساوي الطريقتين في التدريس وبدرجة ثقة 0.99٪ مفترضاً أنّ التوزيع التكراري للمجتمعين طبيعياً.

الحل: لنفترض أنّ μ_1 ، μ_2 يمثلان المتوسطين لمجتمعي دارسي مادة الإحصاء بالطريقتين عندئذٍ باستخدام نفس الطريقة السداسية السابقة وبالتساؤل التالي هل أنّ المجتمعين متكافئان أو هل أنّ طريقتي التدريس العادية والمتلفزة متكافئتان وبعبارة رياضية هل أن $\mu_1 = \mu_2$ أم لا؟ نجد أنّ:

-1 $$H_0: \mu_1 = \mu_2 \text{ أو } \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 \neq \mu_2 \text{ أو } \mu_1 - \mu_2 \neq 0$$

2- تحديد مستوى الأهمية: لدينا درجة الثقة $1-\alpha = 99\%$ ومنه نجد أنّ مستوى الأهمية $\alpha = 0.01$

3- تحديد نوعية التوزيع وإيجاد منطقتي الرفض (الدرجة الحرجة اليمنى واليسرى) بما أنّ حجم العينتين أصغر من 30 لذا نستخدم الدرجة المعيارية التائية وتكون منطقة الرفض هي:

$$t < -t_{\alpha/2} = -1.725 \text{ و } t > t_{\alpha/2} = 1.725$$

حيث: $v = n_1 + n_2 - 2 = 20$.

4- حساب الدرجة المعيارية t_c من العينة:

لحساب كلاً من S_p و t_c نجد من معلومات السؤال بأنّ:

$n_1 = 12$	$\bar{x}_1 = 85$	$S_1 = 4$
$n_2 = 10$	$\bar{x}_2 = 81$	$S_2 = 5$

وبالتعويض في علاقة S_p أو الانحراف التجميعي نجد أنّ:

$$S_p = \sqrt{\frac{(11)(16)+(9)(25)}{v_1 + v_2}} = 4.478$$

وعليه نجد أنّ الدرجة المعيارية التائية هي:

$$t_c = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{(85-81)-0}{4.478\sqrt{\frac{1}{12}+\frac{1}{10}}} = 2.07$$

وذلك بفرض $d_0 = \mu - \mu_0$ وهي تساوي الصفر في مثالنا.

5- التحليل: بالمقارنة مع الخطوة 3 نجد أنّ $t_c = 2.07$ تقع في منطقة الرفض لأنّ:

$$t_c = 2.07 > t_{\alpha/2} = 1.725$$

6- القرار: بناءً على ما سبق نرفض الفرضية H_0 أو تساوي المتوسط في الطريقتين ونقبل الفرضية H_1 أي نقرّ باختلاف الطريقتين للتدريس.

6-3 اختبارات حول التباين (σ^2):

يستخدم اختبار التباين لفحص أو إقرار ما إذا كان تبايني مجتمعين إحصائيين σ_1^2 ، σ_2^2 متكافئاً أو متجانساً أي أنّ تشتت أو ابتعاد مفردات المجتمعين عن متوسطيهما μ_1 ، μ_2 متكافئاً أو متجانساً مثل فحص تجانس نوعين من زيوت المحركات في نقاوتهما أو ارتفاع درجة حرارتهما أو تجانس نوعين من الطلاء الخارجي للسيارات أو البناء من حيث نسبة أو سرعة تغيرهما عند تعرضهما لظروف حرارية أو ظروف معينة أخرى أو شروط صناعية محددة ...إلخ. ونلاحظ في هذا المجال حالتين هما: أ – المقارنة مع تباين محدد أو مفترض σ_0^2 ب– المقارنة بين تبايني مجتمعين مختلفين مثل σ_1^2 ، σ_2^2.

أ – لو افترضنا أن الشرط الصناعي (أو المواصفات العالمية) محددة بـ σ_0^2 مثلاً والمراد هو مقارنة تباين المجتمع σ^2 قيد الدراسة بالمواصفات العالمية فقط أي أنّ $H_0{:}\sigma^2 = \sigma_0^2$ والبديل $H_1{:}\sigma^2 \neq \sigma_0^2$ ففي مثل هذه الحالات الشرطية نستخدم معلمه أو إحصائية مربع كاي مفترضين أنّ التوزيع التكراري للعينات يتقارب مع التوزيع الطبيعي. إنّ إحصائية (معلمة) مربع كاي χ^2 لفحص H_0 أي لفحص تجانس تباين المجتمع قيد الدراسة مع الشرط الصناعي حيث $H_0{:}\sigma^2 = \sigma_0^2$ تعرف بالعلاقة التالية:

$$\chi^2 = \frac{v\ S^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2} \qquad \textbf{(6.3)}$$

حيث n حجم العينة، S^2 تباين العينة و ν درجة الحرية. وكما في السابق نجد أنّ منطقتي الرفض وعلى مستوى الأهمية α هما:

$$\chi^2 > \chi^2_{\alpha/2} \quad و \quad \chi^2 < \chi^2_{1-\alpha/2}$$

أو أنّ منطقة القبول (فترة الثقة) هي:

$$\chi^2_{\alpha/2} < \chi^2 < \chi^2_{1-\alpha/2} \qquad \textbf{(6.4)}$$

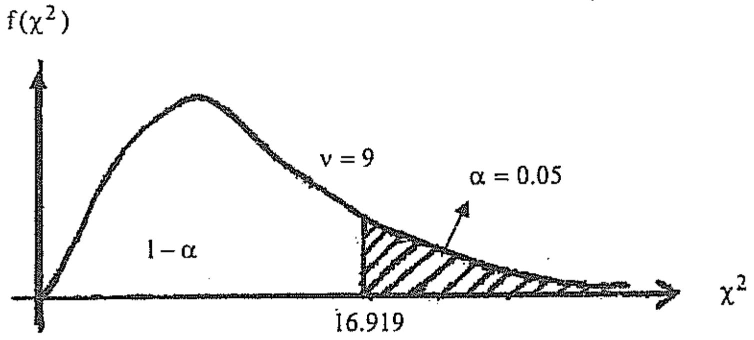
أمّا إذا كانت الفرضية البديلة ذات طرف واحد مثل $\sigma^2 < \sigma_0^2$ نستخدم منطقة الرفض $\chi^2 < \chi^2_{1-\alpha}$ وبالعكس إذا كان $\sigma^2 > \sigma_0^2$ فنستخدم منطقة الرفض $\chi^2 > \chi^2_\alpha$.

مثال (8): تدعي إحدى شركات صناعة بطاريات السيارات بأن الانحراف المعياري لمنتجاتها يساوي 0.9 سنة فإذا أخذت عينة عشوائية من 10 بطاريات للفحص ووجد أنّ انحرافها المعياري هو 1.2 سنة فهل نستنتج أنّ الانحراف المعياري لمنتجات الشركة أكبر من 0.9. استخدم مستوى أهمية $\alpha = 0.05$ للتأكد من ذلك.

الحل: 1- تحديد الفرضيات: $H_0: \sigma^2 = 0.81 \quad H_1: \sigma^2 > 0.81$

2- مستوى الأهمية $\alpha = 0.05$

3- إيجاد الدرجة الحرجة وفقاً لتوزيع مربع كاي ذو درجة الحرية $\nu = 9$ وحيث $\alpha = 0.05$ يتحدد عند النقطة $\chi^2_{0.05,9} = 16.919$ وذلك من الجداول حيث تكون منطقة الرفض $\chi^2 > \chi^2_{(\alpha,9)} = 16.919$



الشكل (6-3)

4- حساب الدرجة المعيارية الكائية المناظرة من العينة:

$$\chi^2 = \frac{\nu\ S^2}{\sigma^2} = \frac{9(1.44)}{0.81} = 16.00$$

5- التحليل: نلاحظ بأنّ الدرجة المعيارية المناظرة χ_c^2 المحسوبة من العينة أصغر من الحرجة أي أنّ: $\chi_c^2 = 16 < \chi^2_{(\alpha,9)} = 16.919$ أي أنها لا تقع في منطقة الرفض بل تقع في منطقة القبول.

6- القرار: نرفض الفرضية H_1 ونقبل H_0 ونستنتج بأنه لا سبب للتشكيك بعدم صحة ادعاء الشركة.

ملاحظة ②: من البديهي أن نستخدم الدرجة المعيارية القياسية (بدلاً من الكائية السابقة) لاختبار $\sigma^2 = \sigma_0^2$ عندما نأخذ حجم العينة العشوائية $n \geq 30$ أي نستخدم الدرجة $z = \frac{S - \sigma_0}{\sigma_0 / \sqrt{2n}}$ والتي تمثل قيمة المتغير العشوائي المعياري الذي يكون توزيع عيناته العشوائية مقارباً للتوزيع الطبيعي.

6-3-1 المقارنة بين تباين مجتمعين:

تتلخص الحالة الثانية باختبار تساوي تبايني مجتمعين إحصائيين ذوي توزيع طبيعي أي أنّ فرضية العدم ستكون $H_0{:}\sigma_1^2 = \sigma_2^2$ مقابل فرضية البديل التي ستمثل بواحدة من الفرضيات الثلاث المعروفة $\sigma_1^2 > \sigma_2^2$ أو $\sigma_1^2 < \sigma_2^2$ أو $\sigma_1^2 \neq \sigma_2^2$ ذات الطرفين. وعليه إذا كانت العينتان العشوائيتان المأخوذتان من المجتمعين مستقلتين وذواتا حجمين n_1 ، n_2 على الترتيب فإنّ الإحصائية المناسبة لفحص H_0 في هذه الحالة هي: $f = \frac{S_1}{S_2}$ والتي تمثل الدرجة المعيارية الفائية لتوزيع فيشر وبدرجتي حرية $v_1 = n_1 - 1$ ، $v_2 = n_2 - 1$ عندما يكون المجتمعان طبيعيي التوزيع. وحيث S_1^2، S_2^2 هما تبايني العينتين العشوائيتين المستقلتين. أمّا مستوى الأهمية ومنطقة الرفض ذات الحجم α المناظرة للفرضيتين ذواتا الطرف الواحد $\sigma_1^2 < \sigma_2^2$ أو $\sigma_1^2 > \sigma_2^2$ فهما على الترتيب تتحددان بالدرجات الفائية الحرجة التالية $f < f_{1-\alpha}(v_1, v_2)$ و $f > f_\alpha(v_1, v_2)$ أمّا فرضية البديل $H_1{:}\sigma_1^2 \neq \sigma_2^2$ ذات الطرفين فتتحدد منطقة الرفض فيها بالمنطقة الحرجة ذات الطرفين التالية:

اليسرى $f < f_{1-\alpha/2}(\nu_1,\nu_2)$ واليمنى $f > f_{\alpha/2}(\nu_1,\nu_2)$

مثال (9): في المثال (7) السابق افترضنا أن تبايني المجتمعين متساويان أي: $\sigma_1^2 = \sigma_2^2$ ولكنهما غير معلومين. فهل هنالك مبرر لمثل هذا الافتراض استخدم مستوى أهمية أو احتمال رفض $\alpha = 0.10$.

الحل: 1- تحديد الفرضيات: $H_0{:}\sigma_1^2 = \sigma_2^2 \quad H_1{:}\sigma_1^2 \neq \sigma_2^2$

2- مستوى الأهمية أو احتمال الرفض $\alpha = 0.10$ ومنه احتمال القبول $1\text{-}\alpha = 0.90$

3- إيجاد الدرجة الفائية الحرجة (منطقة الرفض): إنّ الدرجة الحرجة تحسب وتحدد من الجداول الفائية حيث نجد أنّ: $f_{\alpha/2}(11,9) = 3.11$

$$f_{1-\alpha/2}(11,9) = f_{0.95}(11,9) = \frac{1}{f_{\alpha/2}(9,11)} = 0.34$$

ومنه فإنّ منطقة الرفض هي $f < 0.34$ و $f > 3.11$

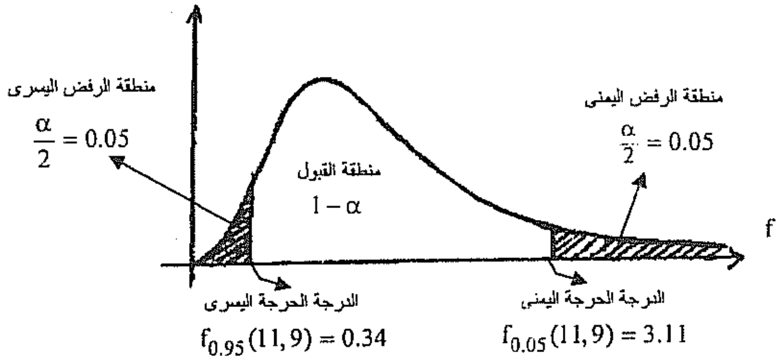
4- حساب الدرجة الفائية للعينتين:

$$f_c = \frac{S_1^2}{S_2^2} = \frac{16}{25} = 0.64$$

5- التحليل: نلاحظ بأنّ الدرجة الفائية للعينتين أو النسبة الفائية أكبر من الدرجة الفائية الحرجة اليسرى وأصغر من الفائية الحرجة اليمنى (راجع الشكل) أي أنّ:

$$0.34 = f_{0.95}(11,9) < f = 0.64 < f_{0.05}(11,9) = 3.11$$

6- القرار: بما أنّ $f = 0.64$ تقع في منطقة القبول لذا نقبل فرضية عدم الاختلاف H_0 ونرفض الفرضية البديلة H_1 وهذا يعني عدم وجود مبرر للشك في عدم تكافؤ أو تجانس المجتمعين أو اختلاف تباينهما.



الشكل (6-4): منحني فيشر بدرجتي حرية $(v_2, v_1) = (11,9)$

وفيما يلي جدولاً يبين الفرضيات والمعلمات الإحصائية المناظرة ومناطق الرفض والقبول:

جدول (6-4): اختبارات التباين σ^2 للمجتمع الإحصائي

شروط ومجالات الرفض	إحصائية الاختبار	الفرضيات
$\chi^2 > \chi^2_{\alpha/2,v}$ $\chi^2 < \chi^2_{1-\alpha/2,v}$	$\chi^2_c = \frac{(n-1)S^2}{\sigma_0^2}$ (لمجتمع واحد)	$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$
$\chi^2 < \chi^2_{1-\alpha,v}$		$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$
$\chi^2 > \chi^2_{\alpha,v}$		$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$
$F > f_{\alpha/2,v_1,v_2}$ $F < f_{1-\alpha/2,v_1,v_2}$	$f_c = \frac{S_1^2}{S_2^2}$	$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$
$F < f_{1-\alpha,v_1,v_2}$	$f_c = \frac{S_2^2}{S_1^2}$	$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 < \sigma_2^2$
$F > f_{\alpha,v_1,v_2}$	$f_c = \frac{S_1^2}{S_2^2}$	$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 > \sigma_2^2$

6-4 اختبارات النسب والاحتمال:

تدخل اختبارات النسب والاحتمال في كثير من المجالات منها نسب العطل أو التلف في المنتجات الصناعية أو نسب التلف في فترات التخزين أو نسب أو احتمالات الفوز والخسارة أو نسب الفشل والنجاح في العمليات والتجارب الكيميائية والدوائية والأمصال الوقائية...إلخ وهنا نلاحظ أيضاً حالتين هما:

أ - الاختبارات الشرطية ونقصد بها اختبار نسب المجتمع الإحصائي قيد الدراسة مع نسبة أخرى مشروطة أو محددة كالنسب العالمية P_0 مثلاً المعترف بها في المجالات الصناعية والاستثمار وبعبارة رياضية $H_0 : P = P_0$ بينما H_1 تأخذ إحدى الفرضيات الثلاث المتعارف عليها $P > P_0$ ، $P < P_0$ أو الفرضية ذات الطرفين $P \neq P_0$ حيث تمثل P هنا معلمة (أو وسيط) التوزيع الثنائي.

ب- اختبار نسب مجتمعين إحصائيين (الاختبارات غير الشرطية) P_2 ، P_1.

في الحالة الأولى نستخدم معلمة التقدير $\hat{P} = \frac{X}{n}$ حيث X يمثل المتغير العشوائي ذو التوزيع الثنائي وحيث تبنى الفكرة على رفض قيم المتغير X البعيدة عن المتوسط الحسابي المشروط $\mu = n P_0$. ومنه لاختبار الفرضية $H_0 : P = P_0$, $H_1 : P < P_0$ وبدرجة ثقة $p = (1-\alpha)\ 100\%$ نوجد العدد الحرج c'_α بحيث أنّ: $P(X \leq c'_\alpha\ ,\ P = P_0) = \sum_{x=0}^{x=c'_\alpha} B(x; n, P_0) \leq \alpha$ ولاختبار الفرضية $H_0 : P = P_0$ و $H_1 : P > P_0$ وبدرجة ثقة $p = (1-\alpha)\ 100\%$ نوجد أيضاً العدد الحرج c_α (الذي يحدد منطقة الرفض) بحيث أنّ: $P(X \geq c_\alpha\ ,\ P = P_0) = \sum_{X=C_\alpha}^{n} B(x; n, P_0) \leq \alpha$ أمّا في حالات اختبار الفرضية ذات الطرفين $H_0 : P = P_0$ و $H_1 : P \neq P_0$ وبدرجة ثقة (منطقة القبول) $p = (1-\alpha)\ 100\%$ أو مستوى أهمية (منطقة الرفض) يساوي α فإنّ الدرجتين الحرجتين تتمثلان بالقيمتين: $X \geq c_{\alpha/2}$, $X \leq c'_{\alpha/2}$ وبما أنّ التوزيع التكراري الثنائي هو من التوزيعات المنفصلة غير المتصلة لذا يتوجب اختيار الأعداد الحرجة c_α ، c'_α ، $c_{\alpha/2}$ ، $c'_{\alpha/2}$ بشكل قريب جداً من منطقة الرفض أمّا أسلوب الاختبار فيتلخص بالخطوات الستة الآنفة الذكر في الفقرة (3-1-6).

مثال (10): يدعي أحد المقاولين بأنه وضع مضخات حرارية في 70٪ من المواقع ضمن تعهده فهل نوافق على هذا الادعاء إذا كان الاستبيان المتخذ من قبل لجنة متخصصة يشير إلى أنّ 8 من كل 15 موقعاً يحتوي على مضخة حرارية وبدرجة ثقة 99٪.

الحل: -1 $H_0{:}P = 0.70$, $H_1{:}P \neq 0.70$

-2 مستوى الأهمية أو احتمال الرفض $\alpha = 1 - 0.99 = 0.10$

-3 المنطقة الحرجة ذات الطرفين (الدرجات الحرجة):

$X \leq 7 = c'_{\alpha/2}$, $X \geq 14 = c_{\alpha/2}$ وذلك من الجداول.

-4 الدرجة المعيارية الثنائية $x = 8$

-5 التحليل: نلاحظ أنّ العدد $x = 8$ يقع في منطقة القبول لأنّ: $c'_{\alpha/2} = 7 \leq 8 \leq c_{\alpha/2} = 14$

-6 نقبل H_0 ونرفض H_1 لعدم وجود سبب أو مبرر مقنع للشك في ادعاء المقاول.

أمّا في الحالة الثانية (ب) غير الشرطية وحيث n (حجم العينة) كبير وأنّ النسبة P_0 تقترب من الصفر أو تقترب من الواحد كثيراً فإنّ التوزيع المستخدم هو توزيع بواسون أمّا في الحالة التي لا تقترب فيها النسبة P_0 من الصفر أوالواحد كثيراً فإنّ التوزيع الطبيعي حيث $\mu = nP_0$ $\sigma^2 = n\,P_0\,q_0$ هو المفضل والدقيق خصوصاً عندما تكون n كبيرة وعليه لاختبار الفرضية $P_1 = P_2$ أو $P = P_0$ فإننا نستخدم المعلمة الإحصائية الطبيعية المعيارية:

$$z = \frac{x - n\ \ P_0}{\sqrt{n\ \ P_0 q_0}} \qquad \textbf{(6.5)}$$

وعليه يتلخص أسلوب الفحص بنفس الخطوات السابقة إلاّ أنّ المناطق الحرجة (مناطق الرفض) تأخذ الشكل التالي:

منطقة يسرى $Z < -z_\alpha$ عندما تكون الفرضية البديلة $P < P_0$ ذات طرف واحد.

منطقة يمنى $Z > z_\alpha$ عندما تكون الفرضية البديلة $P > P_0$ ذات طرف واحد.

منطقة ثنائية $Z < -z_{\alpha/2}, Z > z_{\alpha/2}$ عندما تكون الفرضية البديلة $P = P_1 \neq P_2 = P_0$ ذات طرفين.

مثال (11): للتحقيق في أفضلية دواء جديد A على دواء آخر قديم B يعتقد أنّ نسبة فعاليته 60٪ أخذت عينة من مائة شخص ووجد أنّ 70 منهم قد تم إستشفاؤهم بالدواء A فهل يكفي هذا كدليل وبدرجة ثقة 95٪ بأنّ A أفضل من B.

الحل: 1- $H_0 : P = 0.60 \quad , \quad H_1 : P > 0.60$

2- مستوى الأهمية: $\alpha = 1 - 0.95 = 0.05$

3- المنطقة الحرجة اليمنى (أو منطقة الرفض) تتحدد بالمتباينة

$$Z > z_\alpha = 1.645$$

4- حساب الدرجة المعيارية القياسية:

$$z_c = \frac{x - \mu}{\sigma} = \frac{x - nP_0}{\sqrt{n \quad P_0 q_0}} = \frac{70 - 60}{\sqrt{(100)(0.6)(0.4)}} = 2.04$$

5- التحليل: نلاحظ بأنّ الدرجة المعيارية القياسية z_c أكبر من الدرجة الحرجة الجدولية أي أنّ: $z_c = 2.04 > z_\alpha = 1.645$

6- نرفض H_0 ونقبل H_1 أي نُقِرُّ بأفضلية الدواء الجديد أو أنّ النسبة $P > 0.6$.

6-4-1 اختبار الفرق بين نسبتين P_1 ، P_2 :

غالباً ما تدعو الحاجة أو الدراسة إلى المقارنة بين مجتمعين إحصائيين (أو عدة مجتمعات) من حيث تجانسهما أو تكافؤهما وهذا يعني مقارنة نسب مواصفاتهما قيد الدراسة فيما إذا كانت هذه النسب متساوية أم لا علاوة على مقارنة متوسطاتهما وتباينهما. فمثلاً إذا أردنا مقارنة نسب مبيعات أو أرباح أو خسارة شركتين صناعيتين أو إذا أردنا مقارنة نسب أطباء الأسنان في مجتمعين أو دولتين أو نسب المصابين بمرض معين في منطقتين مختلفتين أو نسب الشفاء نتيجة أخذ لقاح معين...إلخ وعموماً، تصاغ الفرضيات الإحصائية بالشكل التالي:

فرضية العدم $H_0 : P_1 = P_2 = P$ أو $H_0 : P_1 - P_2 = 0$

الفرضية البديلة: $H_1 : P_1 \neq P_2$ أو $H_1 : P_1 > P_2$ أو $H_1 : P_1 < P_2$ أو

$H_1 : P_1 - P_2 \neq 0$ أو $H_1 : P_1 - P_2 > 0$ أو $H_1 : P_1 - P_2 < 0$

حيث P_1 ، P_2 هما نسب الصفة أو الخاصية للمجتمعين الإحصائيين قيد الدراسة.

أمّا الوسيط الإحصائي المستخدم في هذا الاختبار فيؤخذ عن عينتين عشوائيتين مستقلتين ذواتا حجمين كبيرين n_1 ، n_2 من مجتمعين إحصائيين ذوي توزيع ثنائي على الترتيب حيث أنّ

نسب النجاح فيهما هي $\hat{P}_1$ ، $\hat{P}_2$ على الترتيب ومما سبق نلاحظ بأنّ صياغة المعلمة الإحصائية لهذا الاختبار تكون على أساس تقارب التوزيع الثنائي من التوزيع الطبيعي حيث الدرجة المعيارية له هي:

$$z = \frac{(\hat{P}_1 - \hat{P}_2) - d_0}{\sqrt{P_1 q_1 / n_1 + P_2 q_2 / n_2}} = \frac{(\hat{P}_1 - \hat{P}_2) - 0}{\sqrt{Pq \quad (\frac{1}{n_1} + \frac{1}{n_2})}} \qquad (6.6)$$

وحيث $d_0 = P_1 - P_2$ وفي مثل هذه الحالة يتوجب تقدير قيمة P بقيمة النسبة التجميعية $\hat{P}$ (pooling estimate) حيث: $\hat{P} = \frac{x_1 + x_2}{n_1 + n_2}$ وحيث x_1 ، x_2 يمثلان عدد النجاحات أو تكرار النجاحات في كل من العينتين على الترتيب أمّا q فتستبدل بالتقدير $\hat{q} = 1 - \hat{P}$ وعليه نجد أنّ الدرجة المعيارية لاختبار الفرق $P_1 - P_2 = 0$ هي:

$$z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\hat{q} \quad (\frac{1}{n_1} + \frac{1}{n_2})}} \qquad (6.7)$$

أمّا الخطوات الفعلية للاختبار فهي نفس الخطوات الستة المذكورة في اختبار النسب والاحتمال مع حذف إشارة التساوي من الفقرة أو الخطوة الرابعة لتحديد المنطقة الحرجة.

مثال (12): تُصنّع شركتان نوعين مختلفين A، B من محركات توليد الطاقة الكهربائية فإذا تمّ اعتبار 120 من مجموع 200 محرك للشركة الأولى A جيداً بينما تمّ اعتبار 240 من مجموع 500 محرك للشركة الثانية B جيداً أيضاً فهل نستنتج أنّ محركات النوع A أكثر جودة من محركات النوع B وبدرجة ثقة 0.975.

الحل: يمكن التعبير عن التفاوت بالجودة بنسب الإنتاج الجيدة لكل من الشركتين كما يمكن صياغة فرضية العدم أو فرضية عدم التفاوت واللااختلاف في الجودة لكلا المجتمعين الإحصائيين (منتجات الشركتين A، B) كالتالي:

-1 $H_0: P_1 = P_2 \qquad (P_1 - P_2 = d_0 = 0)$

$H_1: P_1 > P_2$

-2 مستوى الأهمية (أو احتمال الرفض): $\alpha = 1 - 0.975 = 0.025$

-3 الدرجة الحرجة المناظرة لمستوى الأهمية α (أو منطقة الرفض):

$$Z > z_\alpha = 1.96$$

4- حساب النسب والاحتمال للعينتين:

$$\hat{P}_1 = \frac{x_1}{n_1} = \frac{120}{200} = 0.60$$

$$\hat{P}_2 = \frac{x_2}{n_2} = \frac{240}{500} = 0.48$$

أمّا النسبة أو الاحتمال التجميعي فهو $\hat{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{360}{700} = 0.51$ وحيث أنّ n_1 ، n_2 كبيرتان جداً لذا نستخدم الدرجة المعيارية القياسية:

$$z_c = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}_1 \hat{q}_1 \quad (\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.60 - 0.48}{\sqrt{(0.51)(0.49)[\frac{1}{200} + \frac{1}{500}]}} = 2.9$$

5- التحليل: نلاحظ بـأنّ الدرجـة المعياريـة $Z = 2.9$ تقـع فـي منطقـة الـرفض لأنّ:

$$z_c = 2.9 > z_\alpha = 1.96$$

6- القرار: نرفض فرضية العدم H_0 ونقبل بالافتراض البديل H_1 أي هنالك فارق كبير فـي الجودة لا يمكن التغاضي عنه لأنّ $P_1 > P_2$ حيث يُفضل في مثـل هـذه الحـالات اسـتخدام محركات توليد الطاقة A بدلاً من محركات توليد الطاقة B لما يترتب على هذا من انخفـاض التكاليف.

ملاحظة ③: عندما تكون فرضية العدم مبنية على أساس اللا تساوي أي وجـود الاخـتلاف والتفاوت في الجودة أو العمـر أو الإنتاجيـة ...إلـخ أي أنّ: $H_0{:}P_1 - P_2 = d_0 \neq 0$ وأنّ الافتراض البديل بخلاف ذلك أيّ $P_1 > P_2$ ، $P_1 < P_2$ أو $P_1 = P_2$ فإنّ الدرجـة المعياريـة المستخدمة في مثل هذه الحالة وعندما يكون الحجم للعينتين كبيراً هي:

$$z_c = \frac{(\hat{P}_1 - \hat{P}_2) - d_0}{\sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}} \qquad \textbf{(6.8)}$$

حيث تمثل Z متغيراً عشوائياً لتوزيع يتقارب كثيراً من التوزيع الطبيعـي القياسـي. بـالعودة للمثال (12) وعلى افتراض أنّ نسب الجودة في إنتاجي الشركتين متقاربة وأنّ الفروقات ليست ذات أهمية تذكر أي أنّ d_0 ضئيل جداً ولنقل أنه يساوي 0.03 نستطيع إعادة الاختبار للمثـال السابق لنجد أنّ:

$$H_0{:}P_1 - P_2 = 0.03 \neq 0 \quad , \quad H_1{:}P_1 - P_2 > 0.03$$

$$z_c = \frac{0.60 - 0.48 - 0.03}{\sqrt{\frac{(0.6)(0.4)}{200} + \frac{(0.48)(0.52)}{500}}} = 2.18$$

ومنها نستنتج رفض الفرضية H_0 نظراً لوقوع $z_c = 2.18$ داخل منطقة الرفض أيّ نــرفض الفرضية القائلة بأنّ الاختلاف ضئيل أو نــرفض القــول بــأنّ $P_1 - P_2 - 0.03 = 0$ أي أنّ مواصفات العينتين ونسب الجودة لهما تشيران إلى فروقات أكبر من ذلك الفــرق بكثيــر ولا تشيران إلى احتمال تساويهما.

جدول (6-5)

اختبارات النسب P (الاحتمال، التكرار النسبي) للمجتمع الإحصائي ومناطق الرفض

شروط ومجالات الرفض	إحصائية الاختبار	الفرضيات
$Z < -z_\alpha$ $Z > z_\alpha$ $Z < -z_{\alpha/2}$, $Z > z_{\alpha/2}$	$z = \frac{x - n P_0}{\sqrt{n \; P_0 q_0}}$ حيث $q_0 = 1 - P_0$	$H_0 : P = P_0$ $H_1 : P < P_0$ أو $P > P_0$ أو $P \neq P_0$ حيث P_0 قيمة أو نسبة محددة مشروطة
$Z < -z_\alpha$ $Z > z_\alpha$ $Z < -z_{\alpha/2}$, $Z > z_{\alpha/2}$	$z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\hat{q} \; (\frac{1}{n_1} + \frac{1}{n_2})}}$ حيث $\hat{P} = 1 - \hat{q} = \frac{x_1 + x_2}{n_1 + n_2}$ تمثل نسبة العينة التجميعية.	$H_0 : P_1 = P_2$ أو $P_1 - P_2 = d_0 = 0$ $H_1 : P_1 < P_2$ أو $P_1 > P_2$ أو $P_1 \neq P_2$
$Z < -z_\alpha$ $Z > z_\alpha$ $Z < -z_{\alpha/2}$, $Z > z_{\alpha/2}$	$z = \frac{(\hat{P}_1 - \hat{P}_2) - d_0}{\sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1} + \frac{\hat{P}_2\hat{q}_2}{n_2}}}$ حيث $\hat{P}_1 = 1 - \hat{q}_1 = \frac{x_1}{n_1}$ $\hat{P}_2 = 1 - \hat{q}_2 = \frac{x_2}{n_2}$ تمثل نسبتي العينتين، n_2،$n_1 \geq 30$ كبيرين جداً	$H_0 : P_1 - P_2 = 0 \pm d_0$ $H_1 : P_1 < P_2$ $P_1 > P_2$ $P_1 = P_2 (d_0 = 0)$

6-5 اختبار التوافق والتجانس (Homogeneity and Goodness of Fit Test):

فيما سبق استعرضنا الاختبارات المتعلقة بمعلمات مفردة للمجتمع الإحصائي مثل μ ، σ^2 ، P والتي قد لا تمثل عنصراً من عناصر المجتمع ذاته. أمّا في هذه الفقرة سندرس الاختبار الذي يتعلق بعناصر المجتمع ذاتها لنرى ما إذا كانت عناصر هذا المجتمع تتمتع بخاصية معينة. واختبار توافق العناصر أو كما يسمى أحياناً اختبار جودة التوفيق المبني على أساس قياس نسب الاختلاف بين تكرارات وقوع العناصر أو الأحداث ضمن عينة عشوائية مأخوذة من مجتمع إحصائي والتكرارات المتوقعة لتلك العناصر محسوبة من توزيع إحصائي معين يشمل المجتمع نفسه. ولتوضيح ذلك نفترض أنّ لدينا قطعة نقود تم رميها 50 مرة ووجد عند انتهاء العملية بأنّ ظهور أو تكرار الصورة هو 31 مرة بينما تكرار أو ظهور الكتابة هو 19 مرة. هذه التكرارات الفعلية الناتجة عن التجربة تسمى التكرارات الملاحظة ويرمز لها بالرمز O_i (observed frequencies) وهي تختلف عن التكرارات المتوقعة التي يرمز لها عادة بالرمز e_i حيث نجد من المثال أنّ التكرارات المتوقعة للصورة هي $P = 25$ والتكرارات المتوقعة للكتابة $q = 1 - P = 25$ بينما نجد أنّ التكرارات الملاحظة للصورة هي 31 وللكتابة هي 19.

جدول التكرارات المتوقعة والملاحظة بتجربة رمي قطعة نقود 50 مرة

	ملاحظة O_i	متوقعة e_i
H	31	25
T	19	25
المجموع	50	50

ولإجراء اختبار التوافق نستخدم المعلمة الإحصائية كاي-مربع χ^2 والتي تعرف بالعلاقة التالية:

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \qquad (6.9)$$

حيث χ_c^2 هي قيمة المتغير العشوائي χ^2 والذي يتقارب توزيعه التكراري مع توزيع مربع كاي. من العلاقة السابقة نلاحظ بأنّ تقارب التكرارات الملاحظة من التكرارات المتوقعة يؤدي إلى قيمة صغيرة لمربع كاي وعليه نستنتج بأنه كلما كانت قيمة مربع كاي صغيرة كان التوافق

والتجانس جيداً وبالعكس كلما كان كبيراً كان التوافق بعيداً أو غير منظور وضعيفاً وعلى هذا الأساس يتم قبول فرضية العدم H_0 إذا كان التوافق جيداً (أي مربع كاي صغيراً) بينما يتم رفض H_0 عندما يكون التوافق أو التجانس ضعيفاً (أو أنّ مربع كاي كبيراً). وبعبارة حسابية إذا كان مربع كاي χ_c^2 أكبر من الدرجة الحرجة لمربع كاي $\chi_{\alpha,v}^2$ أيّ $\chi_c^2 > \chi_{\alpha,v}^2$ أي أنّ القيمة χ_c^2 تقع في منطقة الرفض وفق مستوى أهمية α (أو درجة ثقة $P = 1-\alpha$) ودرجة حرية تعرف بـ $v = (n-1)(m-1)$ حيث n يمثل عدد صفوف (أسطر) جدول الملاحظات أو جدول التوافق بينما m يمثل عدد الأعمدة في الجدول نفسه وفي افتراضنا للتجربة السابقة نجد أنّ درجة الحرية هي: $v = (2-1)(2-1) = 1$ ومن جانب آخر نلاحظ أنّ $n-1$ يمثل عدد العناصر المستقلة في الصف الواحد بينما يمثل $m-1$ عدد العناصر المستقلة في العمود الواحد ومنها نجد أنّ درجات الحرية تمثل عدد العناصر الكلية المستقلة في جدول التوافق أو في التجربة كلها.

مثال (13): لو أردنا اختبار ما إذا كانت العملة المستخدمة في التجربة السابقة متوازنة وغير منحازة مثلاً (أي غير عاطلة أو تالفة) وبدرجة ثقة $P = (1-\alpha) = 95\%$ وهذا يعني أننا نختبر فرضية العدم التالية (بالنسبة لظهور الصورة): 1- $H_0 : O = e$ والفرضية البديلة $H_1 : O > e$

2- تحديد احتمال الرفض: $\alpha = 1 - 0.95 = 0.05$ ودرجة حرية $v = 1$

3- ايجاد الدرجة الحرجة لمربع كاي: فمن الجداول الخاصة نجد أنّ:

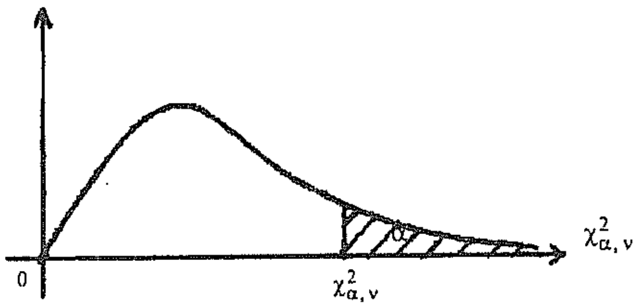
$\chi^2_{(\alpha,v)} = \chi^2_{(0.05,1)} = 3.841$ ومنه فإنّ منطقة الرفض هي $\chi^2 > 3.841$

4- حساب قيمة مربع كاي χ_c^2: وعليه فإنّ قيمة مربع كاي تحسب وفقاً للعلاقة (6.9) كالتالي:

$$\chi_C^2 = \sum_{i=1}^{2} \frac{(O_i - e_i)^2}{e_i} = \frac{(31-25)^2}{25} + \frac{(19-25)^2}{25} = \frac{36}{25} + \frac{36}{25} = \frac{72}{25} = 2.88$$

5- التحليل: بالمقارنة مع الدرجة الحرجة نجد أنّ:

$\chi_C^2 = 2.88 < \chi^2_{(0.05,1)} = 3.841$ أي أنّ $\chi_C^2 = 2.88$ تقع خارج منطقة الرفض أو أنها تقع في منطقة القبول.



الشكل (5-6)

6- القرار: ممّا سبق نستنتج قبول H_0 وعدم رفضها لعدم وجود الدليل الكافي للرفض وأنّ لا مبرر للقول بأنّ هنالك عدم توافق في التكرارات على الرغم من أنّ عدد مرات ظهور الصورة كان $31 > 25$ وعليه نقر بالتوافق بدلاً من عدم التوافق في الجدول السابق.

مثال (14): الجدول التالي يوضح التوزيع التكراري الفئوي المعدّل لأعمار عينة عشوائية متكونة من 40 بطارية للسيارات. اختبر وبدرجة ثقة 95٪ ما إذا كان هذا التوزيع طبيعياً $N(\mu,S)$ وبنفس متوسط العينة μ والانحراف المعياري S لها.

الفترات	(1.45-1.95)	(1.95-2.45)	(2.45-2.95)	(2.95-3.45)	3.45-3.95)	(3.95-4.45)	(4.45-4.95)
الأعمار الملاحظة O_i	2	1	4	15	10	5	3
الأعمار المتوقعة e_i	0.8	2.7	6.9	10.6	10.2	6.0	2.8

الحل: بعد تنظيم الجدول التكراري الفئوي من البيانات الميدانية المأخوذة من العينة ووفقاً للطرق المتعارفة (راجع الفصل الأول) نجد أنّ المتوسط الفئوي لها $\overline{x} = 3.41$ وأنّ الانحراف المعياري للعينة هو $S = 0.703$ ولإجراء الاختبار نلاحظ بأنّ فرضية العدم H_0 هي افتراض عدم وجود اختلاف أو توافق بين هذه البيانات (الأعمار الملاحظة O_i) وبين التوزيع التكراري الطبيعي أمّا الفرضية البديلة فهي H_1 وجود اختلاف بين O_i وقيم التوزيع الطبيعي حيث يتوجب حساب قيم التوزيع الطبيعي Z المناظرة للأعمار O_i أو القيم المتوقعة e_i كالتالي:

في الفئة الأولى نجد أنّ الدرجة المعيارية للحد الأعلى لها هو:

$$z_2 = \frac{1.95-3.41}{0.703} = -2.0768$$

ومن الجدول فإنّ احتمال التكرار للفئة الأولى (مع إهمال الحد الأدنى واعتباره $-\infty$) يساوي: $P(Z < z_2 = -2.0768) = 0.02$ أي أنّ القيمة المتوقعة لتكرار للفئة الأولى $e_i = (0.02)(40) = 0.8$ كما هو موضّح في الجدول أعلاه. بينما نجد في الفئة الرابعة مثلاً وبنفس الأسلوب أنّ الدرجة المعيارية لكل من الحدّين الأدنى والأعلى هي:

$$z_1 = \frac{2.95-3.41}{0.703} = -0.65$$

$$z_2 = \frac{3.45-3.41}{0.703} = 0.06$$

ومن التوزيع التكراري نجد أنّ المساحة تحت المنحني وبين النقطتين أو احتمال وقوع Z بين الدرجتين المعياريتين z_1 ، z_2 هو حسب جدول التوزيع الطبيعي المعياري:

$$P(-0.65 < Z < 0.06) = P(Z < 0.06) - P(Z < -0.65)$$

$$= 0.5239 - 0.2578 = 0.2661$$

ومنه نجد أنّ القيمة المتوقعة لتكرار الفئة الرابعة هي:

$$e_4 = P(z_1 < Z < z_2).n = 0.2661(40) = 10.6$$

أمّا الفئة السابعة فنجد أنّ الدرجة المعيارية للحد الأدنى لها هو:

$$Z_1 = \frac{4.45-3.41}{0.703} = 1.479$$

ومن جدول التوزيع الطبيعي المعياري نجد أنّ:

$$P(Z > z_1 = 1.479) = (0.07)$$

أي أنّ القيمة المتوقعة للفئة السابعة والأخيرة (مع إهمال الحد الأعلى واعتباره $+\infty$) هي:

$$e_7 = P \times 40 = (0.07)(40) = 2.8$$

وبنفس الأسلوب يمكن حساب بقية القيم المتوقعة للفئات الأخرى كما هي مدرجة في الجدول السابق ولحساب مدى التوافق بين الأعمار الملاحظة O_i والأعمار المتوقعة e_i نحسب قيمة مربع كاي للبيانات المجدولة وفقاً للعلاقة (6-9) كالتالي:

$$\chi_c^2 = \frac{(2-0.8)^2}{0.8} + \frac{(1-2.7)^2}{2.7} + \frac{(4-6.9)^2}{6.9}$$

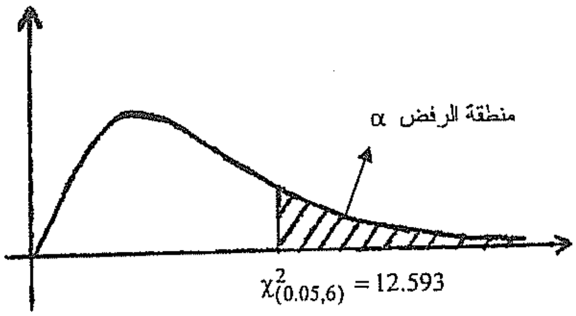
$$+\frac{(15-10.6)^2}{10.6}+\frac{(10-10.2)^2}{10.2}+\frac{(5-6)^2}{6}+\frac{(3-2.8)^2}{2.8}=$$

$$=1.8+1.07+1.219+1.826+0.004+0.167+0.014=6.1$$

وبمقارنـــــة هــــذه النتيجـــة $\chi_c^2 = 6.1$ بالدرجـــة الحرجـــة لتوزيـــع مربـــع كاي $\chi^2_{(\alpha,v)} = \chi^2_{(0.05,6)} = 12.593$ وبدرجة حريـــة $v=(7-1).(2-1)=6$ ومســتوى أهمية $\alpha = 1-0.95 = 0.05$

نجد أنّ: $\chi^2 = 6.1 < \chi^2_{0.05,6)} = 12.592$ أي أنّ: χ_c^2 تقع في منطقة القبول وليس الرفض.

وعليه لا يوجد مبرر لرفض الفرضية H_0 وبالتالي نقر بتقارب أو تســاوي توزيــع أعمــار البطاريات مع التوزيع الطبيعي (N(3.41 0.703



الشكل (6-6)

مثال (15): في محاولة لدراسة نوعية الاستهلاك في ثلاث عواصم ومعرفة ما إذا كان هنالك أي نوع من أنواع التجانس في الاستهلاك بينها قام أحد الاقتصاديين بدراســة عينــة ميــدانياً تتكون من مائة عائلة من كل مدينة ورصد الملاحظات التالية:

جدول البيانات الميدانية O_{ij}

نوع الاستهلاك / المدينة	عالي جداً	عالي	متوسط	ضعيف	حجم العينة
دمشق	42	26	19	13	$M_{10}=100$
طرابلس	55	21	14	10	$M_{20}=100$
عمان	38	30	22	10	$M_{30}=100$
المجاميع	$M_{04}=135$	$M_{03}=77$	$M_{02}=55$	$M_{01}=33$	$M_{00}=300$

الحل: بإتباع أسلوب الخطوات الستة نجد أنّ:

1- فرضية العدم: H_0 نوعية الاستهلاك ليس لها علاقة بمدينة معينة أي أنّ $(O = e)$

الفرضية البديلة: H_1 نوعية الاستهلاك لها علاقة بالمدينة المختارة أي أنّ $(O \neq e)$

2- درجة الثقة $\alpha = 0.05 \Leftarrow P = 95\%$

3- وعليه نجد من الجداول الخاصة بأنّ الدرجة الحرجة لمربع كاي هي:

$\chi^2_{(\alpha, v)} = \chi^2_{(0.05, 6)} = 12.592$ حيث درجة الحرية $v = (4-1).(3-1) = 6$

4- حساب قيمة مربع كاي من العينة χ^2_c: نكوّن جدول القيم المتوقعة e_{ij} لكل حالة من حالات جدول البيانات السابق وذلك باستخدام العلاقة التالية:

$$e_{ij} = \frac{(M_{i0}).(M_{0j})}{M_{00}} \qquad (6.10)$$

ومنها نكون جدول البيانات المتوقعة e_{ij} التالي:

جدول القيم المتوقعة e_{ij}

نوع الاستهلاك / المدينة	عالي جداً	عالي	متوسط	ضعيف	المجموع
دمشق	45	25.7	18.3	11	100
طرابلس	45	25.7	18.3	11	100
عمان	45	25.7	18.3	11	100
المجاميع	135	77	55	33	300

ومن الجدولين يمكننا حساب قيمة مربع كاي الميدانية كالتالي:

O_{ij}	e_{ij}	$(O_{ij} - e_{ij})^2 / e_{ij}$
42	45	0.2
26	25.7	0.004
19	18.3	0.027
13	11	0.364
55	45	2.222
21	25.7	0.860
14	18.3	1.010
10	11	0.091

38	45	1.089
30	25.7	0.719
22	18.3	0.748
10	11	0.091

وبالجمع نجد أنّ: $\chi_c^2 = \sum_{i,j=1}^{3,4} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = 7.425$

5- التحليل والمقارنة: من النتائج أعلاه نجد أنّ قيمة مربع كاي الميدانية المأخوذة عن العينة العشوائية أصغر من القيمة الجدولية الحرجة أي أنّ: $\chi_c^2 = 7.425 < 12.592 = \chi^2_{(0.05, 6)}$ وعليه تقع χ_c^2 في منطقة القبول (وخارج منطقة الرفض) وتشير إلى أنّ هنالك تجانس في الاستهلاك الذي لا يعتمد على نوع المدينة.

6- القرار: وبناءً على المقارنة والتحليل السابق نقبل بالافتراض H_0 القائل بأنّ الاستهلاك ليس له علاقة بمدينة معينة وبدرجة ثقة 95٪ ونرفض البديل H_1 القائل بأنّ الاستهلاك يعتمد على نوع المدينة.

ملاحظة ④: يمكن دمج جدولي البيانات O_{ij} والقيم المتوقعة e_{ij} في جدول واحد كما سنلاحظ في الجدول (6.6).

6-6 اختبار الاستقلالية والترابط (Test for Independence):

يشترك اختبار الاستقلالية أو التجانس لمتغيرين أو أكثر مع اختبار التوافق المذكور سابقاً باستخدام المعلمة أو الإحصائية مربع كاي $\chi^2_{(\alpha,v)}$ المعرّفة سابقاً حيث تحسب هذه الإحصائية من جداول الارتباط والتجانس Contingency tables والتي تعرف بأنها جداول تكرارية ذات $n \times m$ من المواقع (أو الخلايا) التي تضم تكرارات العناصر ضمن صفة أو خاصية معينة وحيث تمثل n عدد العناصر أو الصفوف (الأسطر) في الجدول بينما تمثل m عدد الصفات أو الخواص (الأعمدة) لتلك العناصر، علاوة على القيم الحدودية التي تمثل مجاميع تلك التكرارات عمودياً وأفقياً.

ولحساب القيمة المتوقعة e_i (غير المشاهدة ميدانياً O_i) والمناظرة لكل خلية من خلايا الجدول السابق نلاحظ من تعريف الاستقلالية لعنصرين أو حدثين مثل Y ، X من الأحداث أو

العناصر التالي: $P(X \cap Y) = P(X).P(Y)$ ومنها نجد أنّ التكرار المشترك المتوقع للخلية أو الموقع **ij** عبارة عن حاصل ضرب الاحتمال المشترك والمجموع الكلي للتكرارات أي يحسب من العلاقة التالية:

$$e_{ij} = P(x).P(y).N = \left(\frac{S_i}{N}\right).\left(\frac{S_j}{N}\right).N = \frac{S_i.S_j}{N} \qquad \textbf{(6.11)}$$

حيث N المجموع الكلي للتكرارات في الجدول.

S_i مجموع التكرارات الأفقية في الجدول (i رقم الصف)

S_j مجموع التكرارات الرأسية في الجدول (j رقم العمود)

وعليه فإنّ جدول الاستقلالية والترابط والذي يمكن ترتيبه بصورة مزدوجة معاً يأخذ الشكل التالي:

جدول (6-6) البيانات الميدانية والقيم المتوقعة لها

الصفات المشتركة / العناصر أو المتغيرات	A	B	C	D	المجاميع
X	O_{11} / e_{11}	O_{12} / e_{12}	O_{13} / e_{13}	O_{14} / e_{14}	S_{1-}
Y	O_{21} / e_{21}	O_{22} / e_{22}	O_{23} / e_{23}	O_{24} / e_{24}	S_{2-}
Z	O_{31} / e_{31}	O_{32} / e_{32}	O_{33} / e_{33}	O_{34} / e_{34}	S_{3-}
المجاميع	S_{-1}	S_{-2}	S_{-3}	S_{-4}	S المجموع الكلي

وباستخدام الخطوات الستة الآنفة الذكر لاختبار الفرضيات يمكننا تحديد استقلالية أو عدم استقلالية أي تجانس أو عدم تجانس تلك المتغيرات وفقاً للصفات المعطاة في التجربة الإحصائية.

مثال (16): في تجربة لاختبار مدى العلاقة أو الاستقلالية لثلاثة أنواع A,B,C من محركات السيارات وبين استهلاك نوعين من الوقود L ، H وبدرجـــة ثقـــة أو قبـــول 95٪ رصـــدت المسافات المجدولة التالية بالكيلومترات للسيارات الثلاث:

جدول البيانات الميدانية O_{ij} والقيم المتوقعة لها e_{ij}

أنواع السيارات / أنواع الوقود	A	B	C	S_i
L	182 / 200.9	213 / 209.9	203 / 187.2	598
H	154 / 135.1	138 / 141.1	110 / 125.8	402
S_j	336	351	313	$S=1000$

الحل: بإتباع أسلوب الخطوات الستة نجد أنّ:

1- تحديد الفرضيات: H_0 فرضية العدم: عدم وجود ترابط أي استقلالية أنواع المحركات عن نوعية الوقود المستخدم (أي عدم وجود تأثير لنوعية الوقود على المحركات)

H_1 الفرض البديل: وجود الترابط أو عدم الاستقلالية بين المحركات والوقود (أي وجود تأثير لنوعية الوقود على المحركات)

2- مستوى الأهمية $\alpha = 1 - P = 1 - 0.95 = 0.05$

3- إيجاد الدرجة الحرجة $\chi^2_{(\alpha, v)}$: وبملاحظة أنّ درجة الحرية $v = (2\ \ 1).(3\ \ 1) = 2$ فإنّ الدرجة الحرجة من الجداول الخاصة بمربع كاي هي: $\chi^2_{(0.05, 2)} = 5.991$

4- حساب الدرجة المعيارية لمربع كاي χ^2_c : لحساب هذه المعلمة الإحصائية χ^2_c من الجـــدول السابق يتوجب أولاً حساب e_{ij} كالتالي:

$$e_{11} = \frac{S_{1.} \times S_{.1}}{S} = \frac{598 \times 336}{1000} = 200.9$$

$$e_{23} = \frac{S_{2.} \times S_{.3}}{S} = \frac{402 \times 110}{1000} = 125.8$$

وبنفس الأسلوب يمكن حساب بقية التكرارات المتوقعة e_{ij} (المسافة المتوقعـــة بـــالكيلومترات) حيث نجد أنّ:

$$\chi_c^2=\frac{\sum_{i,j}(O_{ij}-e_{ij})^2}{e_{ij}}=\frac{(182-200.9)^2}{200.9}+\frac{(215-209.9)^2}{209.9}$$

$$+\frac{(203-187.2)^2}{187.2}+\frac{(154-135.1)^2}{135.1}+\frac{(136-141.1)^2}{141.1}+\frac{(110-125.8)^2}{125.8}=8.048$$

5- التحليل والمقارنة: نلاحظ بأنّ الدرجة الكائية $\chi_c^2=8.048$ أكبر من الدرجة الكائية الجدولية $\chi^2_{(0.05,2)}$ أو أن: $\chi_c^2=8.048>\chi^2_{(0.05,2)}=5.991$.

وهذا يعني أنها تقع في منطقة الرفض وليس القبول.

6- القرار: وعليه نرفض فرضية العدم H_0 ونقبل الفرض البديل H_1 أي نقبل بوجود الترابط وعدم الاستقلالية بين محركات السيارات وأنواع الوقود المستخدم.

ملاحظة ⑤: يمكن استخدام مربع كاي لاختبار النسب أو الاحتمالات المتعددة لعدد من المجتمعات الإحصائية حيث نحسب تكرارات الفشل والنجاح أو التالف والجيد لكل عينة من العينات العشوائية المأخوذة من مجتمع معين من المجتمعات الإحصائية المتعددة ومنها يمكن تكوين جدول التجانس والاستقلالية (O_{ij}/e_{ij}) السابق الذكر ومن ثم تطبيق الخطوات الستة المتعارفة للاختبار.

6-7 اختبار الإشارة (Sign Test):

يُعتبر اختبار الإشارة واحداً من الاختبارات البديلة غير الوسيطية بخلاف الاختبارات السابقة التي اعتمدت على افتراض تقارب توزيع التجربة العشوائية من التوزيع الطبيعي خصوصاً عندما يكون حجم العينة كبيراً. ويقصد بغير الوسيطية هنا عدم وجود وسطاء أو إحصائيات محددة عن نوعية التوزيع التكراري للمجتمع الإحصائي قيد الدراسة. ويستخدم اختبار الإشارة عادةً عندما يكون حجم العينة العشوائية $n<30$ والتوزيع التكراري للمجتمع غير طبيعي بالتأكيد ويتلخص هذا الاختبار بوضع إشارة موجبة + بدلاً لكل من متوسطات العينات العشوائية التي تزيد عن μ_0 في فرض العدم $H_0{:}\mu=\mu_0$ بينما توضع إشارة سالبة − بدلاً لكل من متوسطات العينات العشوائية التي تقل عن μ_0 في فرض العدم H_0 وتهمل المتوسطات المساوية لـ μ_0 إذا وجدت ويقلص حجم العينة عادةً عند حصول مثل هذا التساوي. ولو كانت فرضية العدم صحيحة مثلاً $H_0{:}\mu=\mu_0$ والمجتمع متماثل لكان عدد الإشارات الموجبة

والسالبة متساوي تقريباً أمّا إذا اختلف العدد بناءً على محض الصدفة وعدم الانحيـاز فإننـا نرفض فرضية العدم ونقبل الفرضية البديلة $H_1:\mu \neq \mu_0$. والإحصائية المناسبة لهذا الاختبـار هو المتغير العشوائي X الذي يمثل عدد الإشارات الموجبة فـي التجربـة فـإذا كانـت H_0 صحيحة فإنّ احتمالات أن يكون المتوسط الحسابي للعينة موجباً أم سالباً يساوي $P=\frac{1}{2}$ وعلى هذا فإنّ هذه الإحصائية x تكون ذات توزيع تكـراري ثنـائي ذو نسـبة أو معلمـة $P=\frac{1}{2}$ عندما H_0 صحيحة وعليه يتم رفض فرضية العدم H_0 بخلاف ذلك أي عندما $P \neq \frac{1}{2}$ تقريباً.

وعلى سبيل المثال لو أردنا اختبار الفرضية $H_0:\mu=\mu_0$ ضد الفرضية $H_1:\mu<\mu_0$ فإنّ ذلك يعني أنّ $X \leq c'_\alpha$ أي X أصغر من الدرجة الحرجة وحيث c'_α هو العدد الأكبر بحيث أنّ:

$$P(X \leq c'_\alpha ، \mu=\mu_0)=\sum_{x=0}^{C'_\alpha} B(x;n,\frac{1}{2}) \leq \alpha$$

وعندما تكون مساحة منطقة الرفض (مستوى الأهمية) $\alpha=0.05$ مثلاً و $n=15$ نجد مـن جداول التوزيع الثنائي بأنّ: احتمال $X \leq c'_\alpha=3$ عندما $\mu=\mu_0$ هو:

$$P(X \leq 3 \ ، \ \mu=\mu_0)=\sum_{x=0}^{3} B(x;15,\frac{1}{2})=0.0176$$

وكذلك على سبيل المثال عندما $X \leq c'_\alpha=4$:

$$P(X \leq 4 \ ، \ \mu=\mu_0)=\sum_{x=0}^{4} B(x,15,\frac{1}{2})=0.0592$$

وهذا يعني أنّ منطقة الـرفض التـي لا تزيـد عـن $\alpha=0.05$ تتحـدد بالأعـداد $X \leq 3$ وليس $X \leq 4$. وبنفس الأسلوب يتم الاختبـار لـو كانـت الفرضـية البديلـة ذات طـرف واحد $H_1:\mu>\mu_0$ أو ذات طرفين $H_1:\mu \neq \mu_0$ حيـث تؤخـذ منطقـة الـرفض بـدرجتين حرجتين $X \leq c'_{\alpha/2}$ $X \geq c'_{\alpha/2}$ حيث c ، c' عددان حرجان يمكن إيجادهما مـن جـداول التوزيع التكراري الثنائي ذي النسبة أو الاحتمال $P=\frac{1}{2}$ وبعد حساب كل مـن c_α ،c'_α مـن جداول التوزيع الثنائي يمكننا استخدام التوزيع الطبيعي المقارب للتوزيع الثنائي بفرض حجـم العينة $n=20$ حيث يكـون: $\mu=np=20\times\frac{1}{2}=10$ ، $\sigma=\sqrt{npq}=2.236$ والدرجـة المعيارية $z=\frac{x-\mu}{\sqrt{npq}}=\frac{6-10}{2.236}=-1.79$ وبالمقارنة مع الدرجة الحرجة $z_{0.05}=-1.645$

مثلاً نلاحظ وقوع z في المنطقة الحرجة (منطقة الرفض) لذا نرفض H_0 ونقبل بالفرض البديل $\mu < \mu_0$

مثال (17): البيانات التالية تبين زمن الشحن بالساعات لبعض أنواع البطاريات:

1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2, 1.7

استخدم اختبار الإشارة وبمستوى أهمية $\alpha = 0.05$ لاختبار أن هذا الشاحن يعمل بمعدل 1.8 ساعة.

الحل: 1- $H_0: \mu = 1.8 = \mu_0$ $\quad H_1: \mu \neq 1.8$

2- مستوى الأهمية: لدينا $\alpha = 0.05 \Leftarrow \alpha/2 = 0.025$

3- حساب الدرجة الحرجة (منطقة الرفض): القيم الحرجة الجدولية هي $c'_{0.025}$، $c_{0.025}$ ومنه تكون منطقة الرفض $X \leq c'_{0.025}$ ، $X \geq c_{0.025}$، حيث X يمثل عدد الإشارات الموجبة.

4- الحسابات الإحصائية: نستبدل كل عدد يزيد عن 1.8 بالإشارة + وبالعكس بالإشارة – عن كل عدد ينقص عن $\mu_0 = 1.8$ وعليه سنحصل على المتتالية من الإشارات : --+--+--+- حيث $n = 10$ بينما $x = 3$ مع استثناء 1.8 نفسها. 5- التحليل: نلاحظ من جدول التوزيع الثنائي بأنّ $c'_{0.025} = 1$ بينما $c_{0.025} = 9$ حيث نجد أنّ $x = 3$ عدد الإشارات الموجبة تقع ضمن منطقة القبول $1 < 3 < 9$ وليس ضمن منطقة الرفض.

6- القرار: نقبل بفرض العدم $H_0: \mu = 1.8$ ونرفض الافتراض البديل $H_1: \mu \neq 1.8$ وهذا يعني أنّ متوسط عمل الشحن ليس ذا فارق مهم أو كبير.

مثال (18): تدرس إحدى شركات النقل العامة أفضلية استخدام نوعين من العجلات A، B لسياراتها وذلك لتخفيض استهلاك البنزين وعليه اختيرت 12 سيارة نقل وتم تزويدها بالعجلات نوع A واستخدمت لمسافة 50 ألف كيلومتر ومن ثم استبدلت العجلات بالنوع B لنفس السيارات ونفس المسافة فكان استهلاك الوقود في كل من الحالتين كالتالي بالكيلومترات للتر الواحد:

السيارة	1	2	3	4	5	6	7	8	9	10	11	12
العجلات A (μ_1)	4.2	4.7	6.6	7.0	6.7	4.5	5.7	6.0	7.4	4.9	6.1	5.2
العجلات B (μ_2)	4.1	4.9	6.2	6.9	6.8	4.4	5.7	5.8	6.9	4.9	6.0	4.9
$\mu_1 - \mu_2$	0.1	-0.2	0.4	0.1	-0.1	0.1	0	0.2	0.5	0	0.1	0.3

اختبر أفضلية أحد النوعين على الآخر بمستوى أهمية $\alpha = 0.05$.

الحل: لنفترض أنّ μ_1 ، μ_2 يمثلان متوسط الكيلومترات المقطوعة لكل لتر واحد من الوقود وعلى هذا الأساس يتم حل السؤال وفقاً للخطوات الستة المعتادة للاختبارات كالتالي:

1- $H_0: \mu_1 - \mu_2 = 0 \quad (\mu_1 = \mu_2)$

$H_1: \mu_1 - \mu_2 > 0 \quad (\mu_1 > \mu_2)$

2- مستوى الأهمية: لدينا $\alpha = 0.05$

3- منطقة الرفض: ومنه فإنّ الدرجة الحرجة الجدولية $z_\alpha = z_{0.05} = 1.645$ ومنطقة الرفض $Z > z_\alpha = 1.645$

4- حساب الإحصائية χ_c^2 : نلاحظ من الجدول أعلاه بأنّ لدينا إشارتين سالبة وثمانية موجبة ومع إهمال الأصفار نجد أنّ الإحصائية $x = 8$ وعليه باستخدام التوزيع الطبيعي المتقارب مع الثنائي نجد أنّ:

$$\sigma = \sqrt{npq} = 1.581 \quad \text{و} \quad \mu = np = (10)(0.5) = 5$$

ومنه فإنّ الدرجة المعيارية للاختبار $z_c = \dfrac{x-\mu}{\sigma} = \dfrac{8-5}{1.581} = 1.90$.

5- التحليل: نلاحظ بأنّ $z_c = 1.90$ تقع ضمن منطقة الرفض أي أنّ:

$$z_c = 1.90 > z_\alpha = 1.645$$

6- القرار: نرفض H_0 ونقر بأنّ $\mu_1 > \mu_2$ أي أنّ استخدام النوع A من العجلات يساعد على تقليل الوقود والاقتصاد فيه.

6-8 تمارين عامة:

1- يدّعى أن نسبة الشباب من خريجي الجامعات في إحدى المدن الصغيرة هي $P = 0.3$ من ساكنيها ولاختبار هذا الادعاء أخذت عينة من $n = 15$ شاب وكانت النسبة $\frac{2}{7}$ وتم قبول الادعاء فاحسب كلاً من:

أ- احتمال الوقوع في الخطأ من النوع الأول (α).

ب- احتمال الوقوع في الخطأ من النوع الثاني (β).

2- تدعي شركة لصناعة المصابيح الكهربائية بأنّ عمر مصابيحها يتوزع توزيعاً طبيعياً بمتوسط $\mu = 800$ ساعة وبانحراف معياري $\sigma = 40$ ساعة. اختبر كلاً من فرضيتي العدم $H_0{:}\mu = 800$ والقبول $H_1{:}\mu \neq 800$ إذا علمت أنّ عينة عشوائية ذات حجم $n = 30$ قد أخذت من إنتاج الشركة وكان متوسطها $\bar{x} = 788$ مستخدماً مستوى ثقة $P = 1 - \alpha = 0.96$.

3- تدعي شركة لصناعة الخيوط التركيبية، أنّ متوسط قوة التحمل هي 15 باوند بانحراف معياري قدره 0.5. اختبر الفرضية $\mu = 15$ مقابل الفرضية $\mu \neq 15$ وذلك إذا سحبنا عينة عشوائية مكونة من $n = 50$ خيطاً فكان متوسط قوى التحمل 14.8. استخدم مستوى الأهمية $\alpha = 0.01$.

4- إذا كان متوسط المدد اللازمة لتسجيل طالب جامعي هو 50 دقيقة. أتخذ نظام جديد للتسجيل، ثم اختير 12 طالباً وطبق عليهم هذا النظام، فكان المتوسط الجديد 42 بانحراف معياري 11.9. أختبر الفرضية التالية: إنّ متوسط المدد في النظام الجديد هو أقل من 50 دقيقة بمستوى من الأهمية أ- 0.05 ، ب- 0.01

5- عينة مؤلفة من أربع بقرات حلوب كان متوسط إنتاجها اليومي 25 كغ حليب. أختبر الفرضية $H_0{:}\mu = 23$ مقابل الفرضية $H_1{:}\mu > 23$ عند مستوى الأهمية 0.05 وذلك بفرض أنّ الإنتاج يتوزع طبيعياً بانحراف معياري قدره 2.

6- آلة لقطع رقائق معدنية أنتجت في الماضي قطعاً سمكها 0.50 ملم. لتقدير ما إذا كانت الآلة تعمل بصورة جيدة أخذت عينة مؤلفة من $n = 9$ قطع ووجد أنّ معدل سمكها 0.53 ملم

وانحرافها المعياري $S = 0.03$ والمطلوب اختبار الفرضية $H_0{:}\mu = 0.50$ مقابل الفرضية $H_1{:}\mu > 0.50$ عند مستوى الأهمية $\alpha = 0.025$.

7- أخذت عينتان عشوائيتان $n_1 = 25$ ، $\bar{x}_1 = 81$ ، $n_2 = 36$ ، $\bar{x}_2 = 76$ من مجتمعين إحصائيين طبيعيي التوزيع إنحرافهما المعياري $\sigma_1 = 5.2$ $\sigma_2 = 3.4$ على الترتيب. اختبر ما إذا كان متوسطا المجتمعين الإحصائيين متساويان أم لا مستخدماً درجة الثقة $P = 1 - \alpha = 0.94$

8- اختبر الفرضيات: $H_0{:}\mu_1 = \mu_2$ ، $H_1{:}\mu_1 \neq \mu_2$ وبدرجة ثقة 95٪ إذا علمت أنّ قيم العينتين هي: $n_1 = 11$ ، $\bar{x}_1 = 75$ ، $S_1 = 6.1$

$n_2 = 14$ ، $\bar{x}_2 = 60$ ، $S_2 = 5.3$ وذلك على افتراض تساوي تبايني المجتمعين الإحصائيين.

9- البيانات التالية تعطي مقادير التآكل لأنابيب من الصلب المطلية وأنابيب من الصلب غير المطلية داخل أنواع مختلفة من التربة. أحسب فروق أزواج القيم المتناظرة في العينتين واختبر باستخدام توزيع ستيودنت **t** الفرض القائل بأنّ متوسط هذه الفروق يساوي صفراً

غير مطلية	42	37	61	74	55	57	44	55	37	70	52	55
مطلية	39	43	43	52	52	59	40	45	47	62	40	27

10- قطيع من مائة ديك رومي قسما إلى مجموعتين كل مجموعة ذات حجم 50 ديك وتم إطعامهم بنوعين مختلفين من العلف وفي نهاية التجربة تم الحصول على المعلومات التالية بالكيلوات: $S_1 = 2.1$ ، $\bar{x}_1 = 15.2$ ، $S_2 = 2.0$ $\bar{x}_2 = 14.9$ اختبر $\mu_1 = \mu_2$ وأنّ اختلاف العلف ليس ذو أهمية تذكر.

11- ترغب شركة إعلامية في معرفة ما إذا كان متوسط الدخل لأحدى المدن الكبيرة يزداد عن 17500 ليرة أم لا فإذا كانت الشركة ترغب في رفض $H_0{:}\mu = 17500$ وتقبل بالفرضية البديلة $H_1{:}\mu > 17500$ فما هو حجم العينة العشوائية اللازم اختيارها وبدرجة ثقة 0.95 إذا

علمت بأنّ الانحراف المعياري للدخول هو $\sigma = 1000$ ليرة وأنّ المتوسط الحقيقي للدخول هو $\mu = 17600$ ليرة.

12- رامي متمرن أصاب 60٪ من الطلقات المائة الأولى، وفي المائة التالية أصاب 70 منها، هل يمكننا القول أن رمايته قد تحسنت؟ استخدم مستوى الأهمية $\alpha = 0.05$.

13- وضّح ما إذا كان هنالك أي تأثير لملعب الفريق ذاته على فوزه في المباراة أم لا مستخدماً نتائج مباريات كرة القدم لفريق مدينة معينة لمدة 30 عاماً التالية:

مكان المباراة / النتيجة	محلية	خارجية	مجموع النتائج
فوز	$O_1 = 97$	$O_2 = 69$	166
خسارة	$O_3 = 42$	$O_4 = 83$	125
مجموع المباريات	139	152	291

14- إذا كانت الملاحظات الميدانية على عينة متكونة من مائة شخص تبين نوعين من نمط الاستهلاك للسجائر لكل من الذكور والإناث مجدولة بالتالي:

نوع السجائر / الجنس	مالبورو	كنت	المجموع
ذكور	$O_1 = 32$	$O_2 = 28$	$M_{10} = 60$
إناث	$O_3 = 18$	$O_4 = 22$	$M_{20} = 40$
المجموع	$M_{01} = 50$	$M_{02} = 50$	100

فهل هنالك تجانس في نمط الاستهلاك بين الذكور والإناث أم لا؟

الفصل السابع

الانحدار والارتباط البسيط

7-1 – الانحدار الخطي البسيط.

7-2 – الارتباط الخطي البسيط ومعامل بيرسون.

7-3 – ارتباط الرتب ومعامل سبيرمان.

7-4 – معامل كندال لارتباط الرتب (τ).

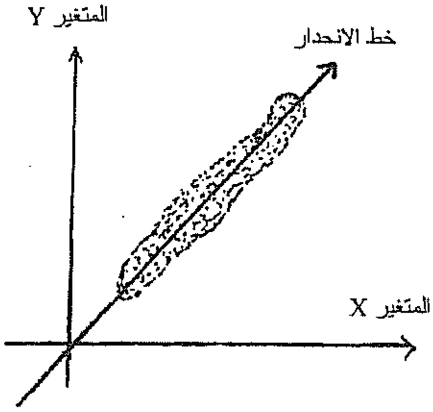
7-5 – تمارين عامة.

الانحدار والارتباط البسيط
(Simple Regression and Correlation)

يُطلق على العلاقة التي تربط المتغير العشوائي Y بالمتغير العشوائي X أسم انحدار المتغيـــر Y على المتغير X بينما يطلق على الرابطة الحسابية بينهما أسم معادلة الانحدار لـ Y على X regression equation كما يطلق على وصف طبيعة تلك العلاقة بين المتغيرين المذكورين اسم تحليل الانحدار regression analysis بينما يطلق على التحقيقات في قوة أو ضعف تلك العلاقة أسم تحليل الارتباط correlation analysis.

ويلاحظ بأنّ كلاً من التسميتين يمثلان ارتباطاً بين المتغيرين إلاّ أن الانحدار يعنـــي حســـاب معادلة الخط البياني بين المتغيرين بينما يعني الارتباط حســـاب قـــوة معامـــل العلاقـــة بـــين المتغيرين (أو المتغيرات المتعددة في حالة الانحدار والارتباط المتعدد غير البسيط).

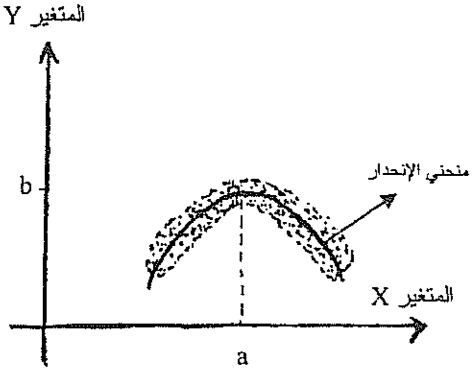
ومن الممكن تحديد وجود الانحدار والارتباط من استخدام الرسوم البيانية التي توضح انتشـــار قيم المتغيرين المتناظرة فمثلاً من خلال الأشكال التالية نلاحظ وجـــود الانحـــدار والارتبـــاط وبصور متعددة.



الشكل (7-1)

الانتشار يوضح وجود علاقة خطية بين قيم المتغيرين، يمكن تمثيلها بالنموذج أو العلاقة الخطية

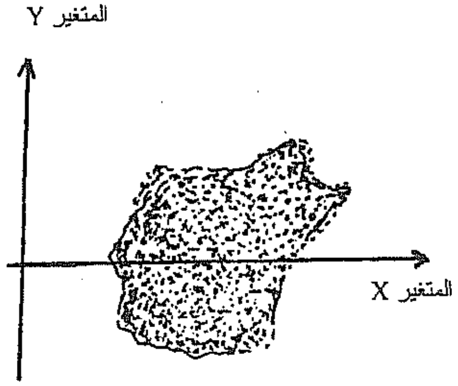
$$y = ax + b$$



الشكل (7-2)

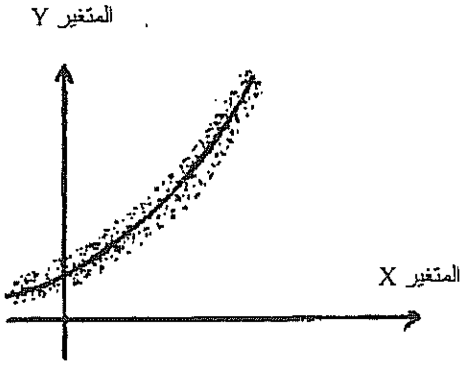
الانتشار يوضح وجود علاقة غير خطية بين قيم المتغيرين يمكن تمثيلها بالنموذج

$$y = (a - x^2) + bx$$



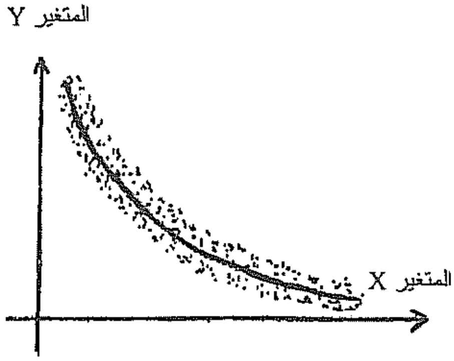
الشكل (7-3)

الانتشار يوضح عدم وجود علاقة بين المتغيرين أي عدم وجود انحدار



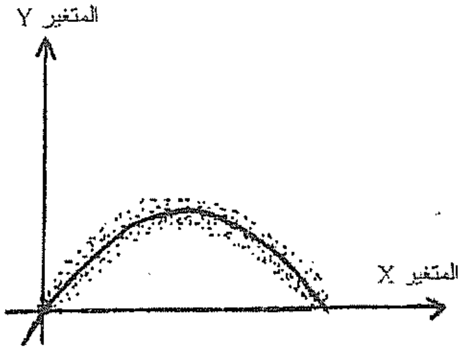
الشكل (7-4)

الانتشار يوضح وجود علاقة غير خطية بين قيم المتغيرين يمكن تمثيلها بالنموذج $y = ae^{\alpha x}$



الشكل (6-7)

الانتشار يوضح وجود علاقة غير خطية عكسية بين قيم المتغيرين يمكن تمثيلها بالنموذج غير الخطي $y = \frac{a}{x}$



الشكل (5-7)

الانتشار يوضح وجود علاقة غير خطية بين المتغيرين يمكن تمثيلها بالنموذج غير الخطي
$y = a\sin\alpha x$

يلاحظ من الشكل (1-7) وجود علاقة لقيم المتغير Y بقيم المتغير X يمكن تمثيلها بشكل خط مستقيم تقريبي يوضح تلك العلاقة حيث يطلق على هذا الخط المستقيم أسم انحدار Y على X في الوقت الذي يمثل انتشار النقاط وجود علاقة خطية بين قيم المتغيرين. أمّا الشكل (2-7) فيوضح انتشار النقاط على وجود علاقة بين قيم المتغيرين X ، Y ولكنها علاقة غير خطية لعدم إمكانية تمثيلها بخط مستقيم بل يمكن تمثيلها بخط منحنٍ نموذجي مثل: $y = (a - x^2) + bx$. وفي الشكل (3-7) نجد أنّ انتشار قيم المتغيرين المتناظرة لا توضح وجود علاقة بين قيم المتغيرين. أمّا في الأشكال (4-7) ، (5-7) ، (6-7) فنلاحظ بأنّ الانتشار يمثل علاقة غير خطية بين المتغيرين يمكن تمثيلها بالنماذج غير الخطية $y = a\ e^{\alpha x}$ ، $y = a\sin\alpha x$ ، $y = \frac{a}{x}$ على الترتيب.

من الأشكال السابقة نجد بأنّ هنالك أربعة أنواع للانحدار هي:

1- الانحدار الخطي البسيط (متغير مستقل واحد)

2- الانحدار الخطي المتعدد أو غير البسيط (عدة متغيرات)

3- الانحدار غير الخطي البسيط (متغير مستقل واحد)

4- الانحدار غير الخطي المتعدد (متعدد المتغيرات)

7-1 الانحدار الخطي البسيط (Simple Regression Line):

من الشكل الانتشاري لقيم المتغيرين يمكننا تمثيل علاقة الانحدار الخطي البسيط بالنموذج الرياضي التالي الذي يمثل معادلة الخط المستقيم:

$$\hat{y} = \alpha + \beta x \qquad (7.1)$$

حيث يمثل الرمز α نقطة تقاطع المستقيم مع المحور y بينما يمثل الرمز β ميل الخط المستقيم أو معامل انحدار النموذج الخطي.

ويتلخص أسلوب حساب كل من المعاملين β, α باستخدام ما يسمى بطريقة المربعات الصغرى التي تعتمد على فرضية الحصول على أصغر قيمة لمربع الخطأ بين القيمة على النموذج المذكور $\hat{y}$ وبين القيمة الواقعية y ولكافة عناصر العينة ويعبر عن ذلك بعبارة رياضية كالتالي:

$$i = 1,\ldots\ldots n \qquad , \qquad E_i^2 = (y_i - \hat{y}_i)^2 \qquad (7.2)$$

حيث E_i يمثل مقدار الخطأ الذي تمّ تربيعه للتخلص من القيمة السالبة للخطأ وبالجمع لكافة قيم i نجد أنّ:

$$G = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2 \qquad (7.3)$$

وباستخدام التفاضل نجد أنّ القيمة الصغرى تحدث عندما تكون المشتقات مساوية للصفر أي نفترض أن المشتق للدالة G يساوي صفراً:

$$\frac{\partial G}{\partial \alpha} = 2\sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))(-1) \equiv 0 \qquad (7.4)$$

$$\frac{\partial G}{\partial \beta} = 2\sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))(-x_i) \equiv 0 \qquad (7.5)$$

وبالتبسيط نجد التالي على الترتيب:

$$\sum y_i = n\alpha + \beta \sum x_i \qquad (7.6)$$

$$\sum x_i y_i = \alpha \sum x_i + \beta \sum x_i^2 \qquad (7.7)$$

ومن المعادلة (7.6) بالقسمة على n نحصل على التالي:

$$\alpha \equiv \overline{y} - \beta \overline{x} \qquad (7.8)$$

وبحل المعادلتين (7.6) و (7.7) آنياً بعد ضرب المعادلة (7.6) بالمقدار $\sum x_i$ والمعادلة (7.7) بالعدد n وإجراء عملية الحذف للحد $n\alpha \sum x_i$ نحصل على التالي:

$$n\sum x_i y_i - (\sum x_i)(\sum y_i) = n\beta(\sum x_i^2) - \beta(\sum x_i)^2 \qquad \textbf{(7.9)}$$

ومنها نجد قيمة ميل الانحدار β كالتالي:

$$\beta \equiv \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \qquad \textbf{(7.10)}$$

ويمكن استخدام المعادلة (7.10) لحساب β بعد حساب عناصرها الأربعة $\sum x_i y_i$ ، $\sum x_i$ ، $\sum y_i$ ، $\sum x_i^2$ ضمن جدول خاص كما يمكن تبسيط المعادلة (7.10) لتأخذ الصورة العامة التالية:

$$\beta \equiv \frac{n\left[\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i\right]}{n\left[\sum x_i^2 - \frac{1}{n}(\sum x_i)^2\right]} \equiv \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

بالضرب بـ $\frac{1}{n-1}$ لكل من البسط والمقام نجد أنّ:

$$\beta = \frac{\frac{1}{n-1}\left[\sum x_i y_i - n\bar{x}\,\bar{y}\right]}{\frac{1}{n-1}\left[\sum x_i^2 - n\bar{x}^2\right]} = \frac{CoV(X,Y)}{S_x^2} \equiv \frac{\text{التباين المزدوج}}{X \text{ التباين لـ}} \qquad \textbf{(7.11)}$$

ومنه:

$$\beta \cong \frac{S_{xy}}{S_x^2} = \frac{\text{تقدير التباين المزدوج للعينة}}{\text{تباين المتغير } X \text{ في العينة}} \qquad \textbf{(7.12)}$$

وبعبارة أخرى نجد أنّ العلاقة (7.11) يمكن أن تكتب على الصورة التالية:

$$\beta = \frac{\frac{1}{n-1}\left[\sum x_i y_i - n\overline{xy}\right]}{\frac{1}{n-1}\left[\sum x_i^2 - n\bar{x}^2\right]} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n-1}\sum(x_i-\bar{x})^2} \qquad \textbf{(7.13)}$$

وذلك بملاحظة أنّ:

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n}(x_i y_i - x_i\bar{y} - y_i\bar{x} + \bar{x}\,\bar{y})$$

$$= \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + n\bar{x}\,\bar{y}$$

$$= \sum_{i=1}^{n} x_i y_i - \bar{y}(n\bar{x}) - \bar{x}(ny) + n\bar{x}\,\bar{y} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}$$

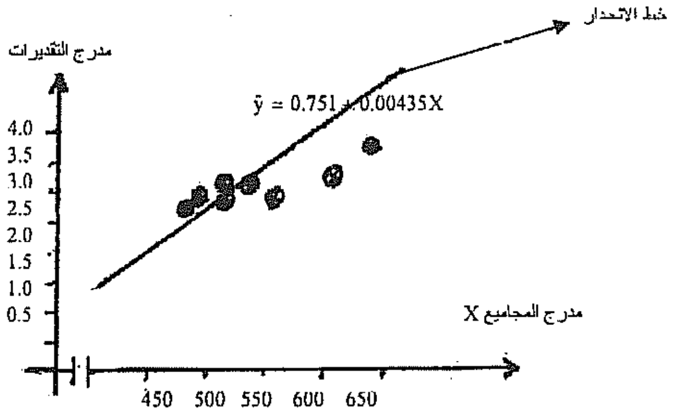
وبالمثل يمكن ملاحظة واستقراء المقام في العلاقة السابقة (7.13).

ملاحظة ①: يمكن استخدام العلاقة (7.13) بدلاً من العلاقة (7.11) لحساب β ميل أو معامل الانحدار للمتغير Y على المتغير X حيث يُسمى تطبيق المعادلة (7.13) بطريقة الانحـــراف عن المتوسط بينما تمثل المعادلة (7.11) الطريقة المباشرة لحساب المعامل β.

مثال (1): الجدول التالي يوضح مجاميع درجات ثمانية طلاب ونقاط التقـــدير لهـــم. أحســـب معادلة انحدار نقاط التقدير على مجاميع الطلبة (إن وجدت).

المجاميع x_i	480	490	510	510	530	550	610	640
التقديرات y_i	2.70	2.90	3.30	2.90	3.10	3.00	3.20	3.70

الحل: نرسم أولاً الشكل الانتشاري على مستوي المحورين y،x بعد تدريجهما.



الشكل (7.7)

الشكل الانتشاري للمثال (1)

يلاحظ من الشكل الانتشاري بأنّ النقاط تتوزع بصورة تقريبية على شكل خـــط مســـتقيم لـــذا نستنتج بأنّ هنالك علاقة خطية بين التقديرات ومجـــاميع الطلبـــة يمكـــن تمثيلهـــا بـــالنموذج الخطي $\hat{y} = \alpha + \beta x$ ولحساب كل من المعاملين α وβ نحسب أولاً β مـــن العلاقـــة (7.13) نكون الجدول التالي: حيث $\bar{x} = 540$, $\bar{y} = 3.10$

x_i المجاميع	y_i التقديرات	$(x_i - \overline{x})$	$(y_i - \overline{y})$	$[(y_i - \overline{y}) \times (x_i - \overline{x})]$	$(x_i - \overline{x})^2$
480	2.70	-60	-0.40	24	3600
490	2.90	-50	-0.20	10	2500
510	3.30	-30	+0.20	-6	900
510	2.90	-30	-0.20	+6	900
350	3.10	-10	0.00	0	100
550	3.00	+10	-0.10	-1	100
610	3.20	+70	+0.10	+7	4900
640	3.70	+100	+0.60	+60	10000
4320	24.80	0	0	100	23000

ومن الجدول نجد أنّ: $\beta = \dfrac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\frac{1}{n-1}\sum(x_i-\overline{x})^2} \equiv \dfrac{100/7}{23000/7} - 0.00435$

وكذلك نجد أنّ: $\alpha = \overline{y} - \beta\overline{x} = 3.10 - 0.00435(540) = 0.751$

وبالتالي فإنّ معادلة الانحدار تأخذ الصورة الخطية التالية: $\hat{y} = 0.751 + 0.00435\ x$

ملاحظة ②: يمكن استخدام نفس الأسلوب عكسياً لإيجاد معادلة انحدار المجاميع على التقديرات أي معادلة انحدار المتغير X على المتغير Y.

ملاحظة ③: يمكن استخدام المعادلة (7.11) لتقدير وحساب كل من معلمتي النموذج الخطي α وβ. (تمرين)

ملاحظة ④: تسمى القيم أو النقاط الواقعة على الخط بأسم القيم التقديرية $\hat{y}$ (Estimated Value) أو القيم المخمنة بينما تسمى القيم المعطاة بأسم القيم المشاهدة observed V. أو القيم الواقعية y. وعلى هذا الأساس يمكن حساب مقدار الخطأ بين القيمة التقديرية والقيمة الواقعية والذي يعرف بالعلاقة التالية:

$$E_i = y_i - \hat{y}_i \qquad \textbf{(7.14)}$$

وبالعودة للمثال السابق يمكن تشكيل جدول الخطأ التالي:

x_i	y_i	$\hat{y}_i = 0.751 + 0.00435\ x_i$	الخطأ المفسر $EE_i = (\hat{y}_i - \bar{y})$	الخطأ غير المفسر الخطأ غير المقبول $NE_i = y_i - \hat{y}_i$
480	2.70	2.839	-0.261	-0.139
490	2.90	2.883	-0.217	0.017
510	3.30	2.970	-0.120	0.330
510	2.90	2.970	-0.120	-0.070
530	3.10	3.057	-0.043	0.0431
550	3.00	3.144	+0.044	-0.144
610	3.20	3.405	+0.305	-0.205
640	3.70	3.535	+0.435	0.165
4320	24.80	24.80	0.023	-0.003

ويستخدم جدول الخطأ السابق لحساب الخطأ الكلي TE حيث $TE_i = EE_i + NE_i$ ومنها نجد أن مجموع مربعات الخطأ يساوي مجموع مربعات الخطأ المفسر + مجموع مربعات الخطأ غير المفسر أو: $SE_i^2 = SEE_i^2 + SNE_i^2$ أو:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \textbf{(7.15)}$$

حيث يكون الانحراف المعياري للخطأ: $\sigma_E = \sqrt{\dfrac{SE_i^2}{n-2}} = \sqrt{\dfrac{\sum(y_i - \hat{y}_i)^2}{n-2}}$

وبقسمة المعادلة (7.15) على مجموع مربعات الخطأ الكلي نجد أنّ:

$$1.00 \equiv \frac{SE_i^2}{SE_i^2} = \frac{SEE_i^2}{SE_i^2} + \frac{SNE_i^2}{SE_i^2}$$

حيث يُطلق على النسبة الأولى أسم مربع معامل التحديد والذي يرمز له بالرمز r أي أنّ:

$$r^2 = \frac{SEE_i^2}{SE_i^2} \rightarrow r = \sqrt{\frac{SEE_i^2}{SE_i^2}} \qquad \textbf{(7.16)}$$

حيث نجد بأنّ: $0 < r^2 < 1$. ويُستخدم المعامل الأخير في تحليل التباين في كل من اختبار t واختبار F.

7-2 الارتباط الخطي البسيط ومعامل بيرسون:

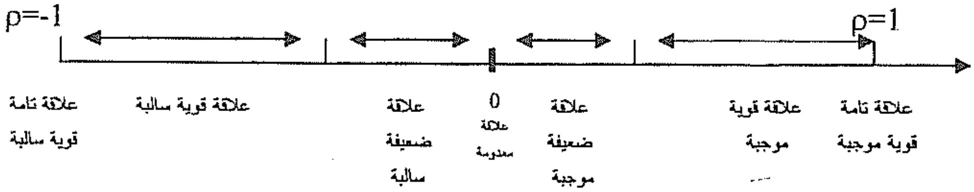
يعرّف الارتباط الخطي البسيط بين متغيرين عشوائيين في مجتمع إحصائي عادة بقوة العلاقة بين المتغيرين المذكورين حيث تقاس هذه القوة بما يسمى بمعامل الارتباط والذي يرمز له بالرمز ρ عادةً. ويعرّف معامل الارتباط الخطي البسيط للمجتمع الإحصائي رياضياً بالعلاقة التالية:

$$\rho = \frac{\text{التباين المزدوج للمتغيرين } X \text{ و } Y}{(\text{الانحراف المعياري للمتغير } X)(\text{الانحراف المعياري للمتغير } Y)} = \frac{CoV(X,Y)}{\sigma x \sigma y} \qquad \textbf{(7.17)}$$

وعادة ما يتم تقدير حساب هذا المعامل باستخدام مفهوم العينات المأخوذة عشوائياً من المجتمع الإحصائي الأم وقيم العناصر x ، y في تلك العينة حيث يحسب من العلاقة التالية:

$$\rho = \frac{S_{xy}}{S_x S_y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad \textbf{(7.18)}$$

حيث يسمى هذا المعامل ρ باسم معامل بيرسون للارتباط البسيط الخطي والذي تنحصر قيمته ضمن الفترة: $-1 \le \rho \le +1$ وعلى ضوء قيم هذا المعامل يتم تصنيف الارتباط بين متغيري العينة Y, X ومن ثم بين المتغيرين في المجتمع الإحصائي بكامله كما في الشكل التالي:



حيث يعبّر عن الارتباط بأنه قوي ايجابي عندما يقترب المعامل ρ من الواحد، وبأنه قوي سلبي عندما يقترب المعامل ρ من 1-، وبأنه ضعيف عندما يقترب المعامل من الصفر.

ويمكن تبسيط العلاقة الحسابية (7.18) التي تعتمد على حساب الانحرافات عن المتوسط الحسابي للمتغيرين Y, X إلى علاقة تعتمد على البيانات المعطاة مباشرة كالتالي:

$$\rho = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}\sqrt{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2}} \qquad (7.19)$$

حيث نحصل على نفس النتيجة. ومن الواضح في كـل مـن العلاقتـين (7.18) و (7.19) ضرورة الترتيب الجدولي لعناصر معامل الارتباط الخطي ρ والذي يسمى بمعامل بيرسـون للارتباط الخطي البسيط أيضاً.

ملاحظة ⑤: يستخدم معامل بيرسون للارتباط الخطي البسيط لحساب قوة أو درجة الارتبـاط في العينة ومن ثم تقدير قوة أو درجة الارتباط في المجتمع الإحصائي بغض النظر عن معادلة الانحدار بين المتغيرين وحتى من دون اللجوء إلى حسابها.

ملاحظة ⑥: إنّ وجود ارتباط بين متغيرين عشوائيين مثل Y, X لا يعني على الإطلاق بـأنّ أحدهما سبب أو علة للآخر لأننا وببساطة يمكن أن نرتب بعضاً من الأرقام ليكـون معامـل الارتباط قوياً في حين أنّ الأرقام المرتبة لا تمثل أي سببية أو عليّة لأحـد المتغيـرين علـى الآخر.

مثال (2): أحسب معامل بيرسون للارتباط بين المتغيرين Y, X التاليين:

x	1	3	4	6	8	10
y	2	6	7	10	15	20

الحل: نكوّن الجدول التالي:

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2	2	1	4
3	6	18	9	36
4	7	28	16	49
6	10	60	36	100
8	15	120	64	225
10	20	200	100	400
32	60	428	226	814

وبتطبيق العلاقة (7.19) نجد أنّ:
$$\rho = \frac{428 - \frac{1}{6}(32)(60)}{\sqrt{226 - \frac{1}{6}(32)^2}\sqrt{814 - \frac{1}{6}(60)^2}} = 0.992$$

ومنه نستنتج بأنّ الارتباط قوي جداً و إيجابي بين المتغيرين.

7-3 ارتباط الرتب ومعامل سبيرمان (Rank Correlation):

يُستخدم مفهوم الترتيب في الحالات غير الكمية التي تشمل الصفات المعنوية حيث تكون الرتب بديلاً كمياً يمكن التعامل معه. ويسمى معامل الارتباط لهذه الحالات باسم معامل ارتباط الرتب أو معامل سبيرمان والذي يُعرّف بالعلاقة التالية:

$$r = 1 - \frac{2\sum_{i=1}^{n}(R_{x_i} - R_{y_i})^2}{Max\sum_{i=1}^{n}(R_{x_i} - R_{y_i})^2} = 1 - \frac{6\sum_{i=1}^{n}(R_{x_i} - R_{y_i})^2}{n(n^2-1)} \qquad \textbf{(7.20)}$$

حيث: R_{x_i} يمثل رتبة المتغير x_i و R_{y_i} يمثل رتبة المتغير y_i

n يمثل عدد الأزواج المرتبة $(x ، y)$

ويتلخص أسلوب حساب معامل ارتباط سبيرمان بترتيب الصفات بصورة متسلسلة حسب معيار معين تصاعدياً أو تنازلياً ومن ثم يتم إعطاء الرتب وفقاً للتسلسل المذكور. ومن الرتب المتناظرة يمكننا حساب الفروق بينها $R_{x_i} - R_{y_i}$ ولكافة الرتب.

ملاحظة ⑦: عندما تتكرر صفة من الصفات فإنّ الرتبة التي تُعتمد في مثل هذه الحالة هي متوسط التسلسل لهذه الصفة (أي مجموع التسلسل للصفة المتكررة على عدد تكرارها).

فلو افترضنا أنّ الصفة المكررة (جـ) أخذت التسلسلات التالية 6 ، 9 ، 12 أي تكررت ثلاث مرّات فإنّ رتبة الصفة (جـ) هي $R = \frac{27}{3} = 9$ للحالات المكررة الثلاث.

مثال (3): في أحد المنافسات تم استعراض 10 لوحات فنية لتقييمها من قبل لجنتي تحكيم فإذا كانت الرتب الممنوحة للوحات العشرة من قبل اللجنتين هي كالتالي:

اللجنة الأولى	1	2	3	4	5	6	7	8	9	10
اللجنة الثانية	4	6	1	2	3	7	8	10	5	9

فأحسب معامل سبيرمان للارتباط وحدد نوعيته.

الحل: نرتب جدول العناصر لمعامل سبيرمان كالتالي:

اللوحات	r_1 لجنة أولى	r_2 لجنة ثانية	$r_1 - r_2$	$(r_1 - r_2)^2$
A	1	4	-3	9
B	2	6	-4	16
C	3	1	2	4
D	4	2	2	4
E	5	3	2	4
F	6	7	-1	1
G	7	8	-1	1
H	8	10	-2	4
I	9	5	4	16
J	10	9	1	1
				المجموع = 60

ومنه فإنّ:

$$r = 1 - \frac{6\sum_{i=1}^{10}(r_1 - r_2)^2}{n(n^2-1)} = 1 - \frac{6(60)}{10 \times 99} = 1 - \frac{60}{165} = \frac{105}{165} = 0.636$$

ويلاحظ بأنّ هنالك ارتباطاً قوياً ايجابياً في التقييم بين اللجنتين في مضمار العينـــة المـــأخوذة.

ولأجل تعميم النتيجة على المجتمع الإحصائي لا بدّ من اســتخدام جـــدول التوزيـــع التـــائي (t-distribution) لمعرفة مدى دلالة الرقم $r = 0.636$ على المجتمع الإحصائي حيث:

$$t = \frac{r_s}{S_s} = \frac{r_s}{\sqrt{\frac{1 - r_s^2}{n-2}}} = \frac{0.636}{\sqrt{\frac{1 - 0.404}{10 - 2}}} = \frac{0.636}{0.273} = 2.3297$$

حيث $(n-2)$ تمثل درجات الحرية.

وباستخدام جدول توزيـــع t نجـــد أنّ الـــرقم $2.33 \approx 2.3297$ أي $P(t < 2.33)$ ذو دلالـــة إحصائية عند المستوى 0.01 وبهذا نستطيع الجزم بوجود ارتباط بين الرتب وعليـــه نـــرفض الافتراض H_0 القائل بعدم وجود أي ارتباط.

مثال (4): في إحدى الدراسات الاجتماعية لمعرفة مقدار العلاقة بين الحالة العلمية لرب الأسرة وتأثيره على المستوى الاقتصادي لها استُحصلت البيانات التالية:

الحالة العلمية لرب الأسرة	ابتدائية	أمي	عالية	متوسطة	أمي	يقرا ويكتب	أمي	يقرا فقط	ثانوية	جامعية
المستوى الاقتصادي للأسرة	متوسط	معدمة	غنية	فقيرة	فقيرة	فقيرة	متوسط	فقيرة	متوسط الحال	غنية

فأحسب معامل الارتباط للصفات في الدراسة أعلاه.

الحل: نُرتّب أولاً الحالات العلمية لرب الأسرة بصورة متسلسلة وتصاعدية وفقاً للدرجة العلمية ومن ثم نرتب المستويات الاقتصادية بصورة متسلسلة ومنفصلة عن الترتيب الأول لتكوين أزواج الرتب الثنائية (x_i, y_i) ومن ثم نوجد مجموع فروقها كما في الجدول التالي حيث تأخذ العينات نفس تسلسلها ما لم تكن مكررة حيث نلاحظ بأنّ رتبة الحالة العلمية (أمـي) مكـررة ثلاث مرّات وعليه فإنّ رتبتها تساوي مجموع التسلسلات على التكرار أي أنّ:

$$r\ (\text{أمي})\ -\frac{1+2+3}{3}=2$$

$$r\ (\text{فقيرة})\ =\frac{2+3+4+5}{4}=\frac{14}{4}=3.5$$

$$r\ (\text{متوسطة})\ =\frac{6+7+8}{3}=\frac{21}{3}=7$$

التسلسل	الحالة العلمية	الرتبة	المستوى الاقتصادي للأسرة	الرتبة	الفروق d_i للرتب	مربع الفروق للرتب d_i^2
1	أمي	2	معدمة (1)	1	+1	1
2	أمي	2	فقيرة (2)	3.5	-1.5	2.25
3	أمي	2	متوسط الحال (6)	7	-5	25
4	يقرأ فقط	4	فقيرة (3)	3.5	+0.5	0.25
5	يقرأ ويكتب	5	فقيرة (4)	3.5	+1.5	2.25
6	شهادة ابتدائية	6	متوسطة الحال (7)	7	-1	1
7	شهادة متوسطة	7	فقيرة (5)	3.5	3.5	12.25
8	شهادة ثانوية	8	متوسطة الحال (8)	7	+1	1
9	شهادة جامعية	9	غنية (9)	9.5	-0.5	0.25
10	شهادة عالية	10	غنية (10)	9.5	+0.5	0.25

ومنه نجد أنّ معامل سبيرمان للرتب:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{273}{990} = 0.724$$

والقيمة الأخيرة تدلل على وجود علاقة قوية موجبة بين الحالة العلمية لرب الأسرة والمستوى الاقتصادي لها.

7-4 معامل كندال لارتباط الرتب (τ) (Kendall's corr.co):

يعتمد معامل كندال في حساب الارتباط على تحديد مؤشر تبدل رتبة العنصر بالمقارنة مع عنصر آخر حيث تُعطى إشارة (+1) عندما تكون الرتبة موافقة للترتيب الأول المفترض (تصاعدياً أو تنازلياً) بينما تُعطي إشارة (-1) عندما تكون الرتبة بعكس الترتيب الأول وعلى هذا الأساس نجد أنّ معامل كندال لا يعتمد مباشرةً على الرتب ذاتها بل على مؤشراتها المفترضة بعكس معامل سبيرمان الذي يعتمد على رتب الصفات ذاتها.

يعرّف معامل كندال الذي يرمز له بالرمز τ بالعلاقة التالية:

$$\tau = \frac{\text{مجموع المؤشرات المحسوبة}}{\text{أكبر مجموع محتمل للمؤشرات}} = \frac{\sum \pm 1}{C_n^2}$$

حيث n يمثل عدد الرتب بينما العدد 2 يرمز لزوج الرتب المطلوب تأشيرها أو تحديد مؤشرها من مجموع n رتبة (حيث أنّ أكبر مجموع محتمل يحدث عندما تتوافق الرتب تماماً)

وحيث: $C_n^2 = \frac{n!}{2!(n-2)!}$.

مثال (5): في المثال السابق للتحكيم في استعراض لوحات الرسم العشرة نجد التالي:

رتب اللجنة الثانية	رتب اللجنة الأولي	اللوحات
4	1	A
6	2	B
1	3	C
2	4	D
3	5	E
7	6	F
8	7	G
10	8	H
5	9	I
9	10	J

المقارنة المزدوجة بين رتب اللوحات			المؤشر		
(1)	(2)	(3)	(1)	(2)	(3)
AB	CD	FG	+1	+1	+1
AC	CE	FH	-1	+1	+1
AD	CF	FI	-1	+1	-1
AE	CG	FJ	-1	+1	+1
AF	CH	GH	+1	+1	+1
AG	CI	GI	+1	+1	-1
AH	CJ	GJ	+1	+1	+1
AI	DE	HI	+1	+1	-1
AJ	DF	HJ	+1	+1	-1
BC	DG	IJ	-1	+1	+1
BD	DH		-1	+1	
BE	DI		-1	+1	
BF	DJ		+1	+1	
BG	EF		+1	+1	
BH	EG		+1	+1	
BI	EH		-1	+1	
BJ	EI		+1	+1	
	EJ			+1	

حيث نجد أنّ:

$$\tau = \frac{\sum \pm 1}{C_{10}^{2}} = \frac{34-11}{45} = \frac{23}{45} = 0.511$$

ويلاحظ من النتيجة $(\tau = 0.511)$ بأنّ هنالك ارتباطاً إيجابياً قوياً نسبياً بين قراري اللجنتين يتوافق مع معامل سبيرمان للارتباط $r_S = 0.636$ ويلاحظ بأنّ قيمة معامل كندال τ تنحصر أيضاً في الفترة: $-1 \le \tau \le +1$ فعندما تكون الرتب متوافقة تماماً فإنها تأخذ المؤشر $(+1)$ جميعاً وبالتالي نجد أنّ: $\tau = \frac{+C_n^2}{C_n^2} = +1$ وبالعكس إذا كانت الرتب مختلفة تماماً فإنها تأخذ المؤشر (-1) جميعاً وبالتالي نجد أنّ $\tau = \frac{-C_n^2}{C_n^2} = -1$ ومن المهم جداً معرفة دلالة الرقم المحسوب $(\tau = 0.511)$ مثلاً ضمن المجتمع الإحصائي وليس للعينة فقط وبهذا يستخدم التوزيع الطبيعي لحساب هذه الدلالة بعد تحويل المعامل إلى الدرجة المعيارية كالتالي:

الدرجة المعيارية لمعامل كندال للرتب: $z_k = \frac{\tau - \mu_k}{S_k}$ حيث k ترمز إلى كندال و S_k يرمز إلى الانحراف المعياري لكندال حيث: $S_k = \sqrt{\frac{2\ \ (2n+5)}{9\ \ n(n-1)}}$

وعند افتراض عدم وجود علاقة أو رابطة (تحت الفرضية الصفرية H_0) أي أنّ: $\mu_k \equiv 0$
نجد من المثال السابق بأنّ:

$$S_k = \sqrt{\frac{2\ \ (20+5)}{90(9)}} = \sqrt{\frac{50}{810}} = \sqrt{0.061728395} = 0.2485 \cong 0.249$$

ومنها نجد أنّ الدرجة المعيارية: $z_k = \frac{\tau}{S_k} = \frac{0.511}{0.249} = 2.052$

وباستخدام جداول التوزيع الطبيعي وبما أنّ: $P(Z \le z_k) = P(Z \le 2.052) = 0.9798$

وأنّ $P(Z > z_k) = P(Z > 2.052) = 1 - 0.9798 = 0.0202 > 0.01$

ومن الاحتمال الكبير $(\alpha = 0.0202)$ نستنتج بأنّ معامل الارتباط ذا دلالة وبهذا يمكن رفض الافتراض الصفري H_0 القائل بعدم وجود علاقة ارتباط للرتب.

7-5 تمارين عامة:

1- البيانات التالية تمثل الطول x والوزن y لاثني عشر شخصاً

أ- أرسم الشكل الانتشاري واحسب معادلة انحدار Y على X.

ب- أحسب قيمة معامل الارتباط لبيرسون

x	65	73	70	68	66	69	75	70	64	72	65	71
y	124	184	161	164	140	154	210	164	126	172	133	150

2- البيانات التالية توضح مقاومة الشد وصلابة الألمنيوم المسبوك. أحسب معادلة انحدار مقاومة الشد Y على صلابة الألمنيوم X وأرسم الشكل الانتشاري:

T.S(y)	293	349	368	301	340	308	354	313	322	334	377	247
H(x)	53	70	84	55	78	64	71	53	82	67	70	56

3- البيانات التالية تبين مقدار ضغط الغاز لحجوم مختلفة:

$V : cm^3$	50	60	70	90	100
$P : kg/cm^2$	64.7	51.3	40.5	25.9	7.8

قدر مقدار الضغط P عندما $V = 80cm^3$ إذا علمت أنّ $PV^{\gamma} = C$ حيث C،γ ثوابت

أحسب تقدير كل من C،γ بطريقة المربعات الصغرى.

4- البيانات التالية توضح مسافة التوقف وفقاً لسرعات مختلفة في تجربة على موديل جديد للسيارات.

V السرعة كيلومترات بالساعة	35	50	65	80	95	110
d مسافة التوقف بالأمتار	16	26	41	62	88	119

والمطلوب: أحسب معامل الارتباط لبيرسون.

5- أحسب الميل ونقطة القطع مع المحور y واكتب معادلة انحدار المتغير Y على المتغير X من البيانات التالية:

y	3	19	18	22	23
x	-2	-1	0	1	2

مستخدماً طريقة المربعات الصغرى.

6- إذا كانت معادلة انحدار المتغير Y على X هي: $y = a_0 + a_1x$ ومعادلة انحدار المتغير X على Y هي $x = b_0 + b_1y$ فأثبت أنّ $a_1b_1 = r^2$ حيث r معامل الارتباط بينهما.

7- أحسب معادلة انحدار متوسط التكلفة y لإنتاج عدد من السلع x شهرياً باستخدام الجدول التالي:

متوسط التكلفة	3	9	40	15	33	45	20
كمية الإنتاج شهرياً	15	10	18	10	15	20	13

ثم أحسب جدول الخطأ وأوجد معامل التحديد r^2

8- إذا كان خط الانحدار للإنفاق السنوي على المواد الغذائية لأسرة معينة يتمثل بالمعادلة الخطية التالية $y = 300 + 0.2x$ حيث x هو الدخل السنوي للأسرة بينما y مقدار الإنفاق السنوي والمطلوب:

أ – أرسم خط انحدار الإنفاق على الدخل للأسرة.

ب– ما هو إنفاق الأسرة (المصروف السنوي) إذا كان دخلها 30000 ليرة ، 16000 ليرة.

جـ – لماذا تهمل العلاقة السابقة عندما يكون الدخل صفراً.

9- في أحد البيانات المجدولة وجد أنّ:

$$\sum y_i^2 = 160 \ ، \ \sum x_i^2 = 10 \ ، \ \sum x_iy_i = -30 \ ، \ \sum y_i = 30 \ ، \ \sum x_i = 15 \ ، \ N = 30$$

فأحسب كلاً من معادلتي انحدار Y على X و X على Y.

10- أدرس العلاقة بين مستوى الطلاب في مادتي الرياضيات والإحصاء وذلك باستخدام التقديرات التالية لأثنا عشر طالباً:

F	F	C	C	C	D^+	D^+	C^+	A	B	B	C	تقديرات الإحصاء
F	F	D	D	D	C^+	C	B	A	A	B	C	تقديرات الرياضيات

11- في دراسة عن أولويات الطلبة للانتماء إلى الأقسام الخمسة كهرباء E1، تصميم وإنتاج P، إلكترون E2، غزل ونسيج D، ميكانيك M في إحدى الكليات لوحظ الترتيبان التاليان:

الترتيب الأول	E1	E2	P	M	D
الترتيب الثاني	E2	M	E1	P	D

أحسب معامل ارتباط الرتب لكندال.

الفصل الثامن

الأرقام القياسية والسلاسل الزمنية

8-1 – مكونات السلاسل الزمنية.

8-2 – نماذج السلاسل الزمنية.

8-3 – التوجه العام (أو النمط) الخطي للسلسلة.

8-4 – التوجه العام T (أو النمط) غير الخطي.

8-5 – طريقة المتوسطات لتعديل السلسلة الزمنية.

8-6 – حساب وتقدير مركبة المؤثرات الموسمية.

8-7 – حساب وتقدير مركبة المؤثرات الدورية.

8-8 – الأرقام القياسية.

8-9 – تمارين عامة.

الأرقام القياسية والسلاسل الزمنية
(Index numbers and Time Series)

تُعرف السلاسل الزمنية على أنها دوال ذات متغير مستقل واحد هو الزمن t والذي يُقاس عادة لفترات طويلة كالسنين أو الأشهر أو الأسابيع وعلى أساس هذا التعريف نلاحظ أنّ أغلب البيانات التجارية والاقتصادية وحتى الإنتاجية تأخذ صورة أو شكل السلاسل الزمنية. فالمبيعات الشهرية لمؤسسة إنتاجية سلعية أو إجمالي الناتج القومي السنوي (Gross National product G.N.P)، أو بيانات الاستهلاك وتغيرات معدلات الأسعار مثل أسعار النفط تشكل أمثلة للسلاسل الزمنية. وعادة ما يتم التركيز على نوعين من المسائل في تحليل السلاسل الزمنية هما:

1- استقراء وتقدير المكونات التي تحدد نوعية السلسلة الزمنية.

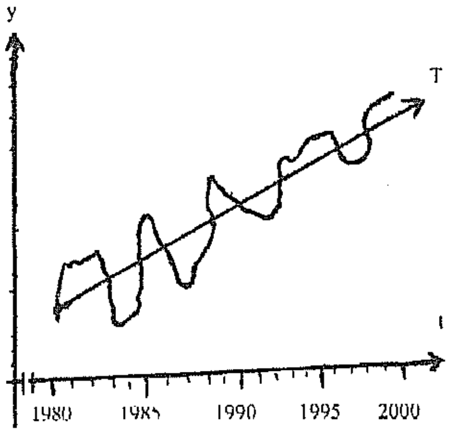
2- استخدام تلك التقديرات والحسابات للتنبؤ بسلوك السلسلة للمراحل المستقبلية والذي يتم على أثره اتخاذ القرار المناسب وفي هذا المجال تؤدي السلاسل الزمنية دوراً هاماً.

ومن جملة التطبيقات أو الخدمات الهامة التي تقدمها السلاسل الزمنية هو تقدير معدلات الفائدة السنوية ومعدلات البطالة ونسب زيادة التكلفة وغلاء المعيشة...إلخ. ولربما تستفيد الشركات الصناعية من هذه السلاسل لمعرفة مقدار الطلب على منتجاتها من قبل الأفراد أو الأسواق خلال الأعوام المقبلة مثل شركات السيارات والملابس...إلخ.

8-1 مكونات السلاسل الزمنية:

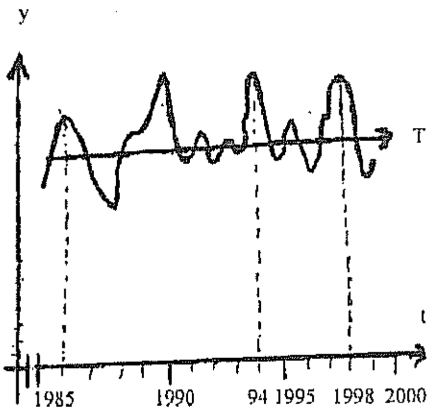
تتكون السلاسل الزمنية من أربعة مكونات هي:

1- النمط أو التوجه العام (T): والذي يمثل التبدلات والتغيرات على طول فترة الدراسة الكاملة للسلسلة وعلى نفس الوتيرة فمثلاً نجد أنّ معدلات الناتج القومي قد تتغير بنسبة 2٪ إلى 3٪ سنوياً وعلى مدى عشرين عاماً وهذا التوجه العام والذي يرمز لـه بـالرمز T(Trend) على مدى العشرين عاماً قد يدفع الحكومة مثلاً إلى تقليل ساعات العمل أو زيادة مستوى الأسعار حسب ظروف السوق. ويمكن تمثيل T بالشكل التالي:



الشكل (8-1-a)

يوضح خط التوجه العام للسلسلة على مدى 20 عاماً



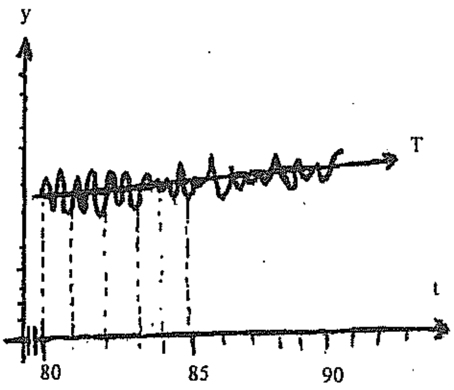
الشكل (8-1-b)

يوضح التبدلات الدورية كل أربع سنوات

2- التغيرات الموسمية (S): وتمثل التبدلات والتغيرات المتكررة خلال فترة السنة الواحدة أو فترة محددة الطول كفترتي الصيف والشتاء. حيث نجد أنّ للكثير من السلاسل مكونات موسمية أو فصلية مثل إنتاج الفواكه في فصل معين أو العمالة في الصيف أو كمية الأمطار في شهر كانون أو الطلب على بعض السلع كسلع السباحة في الصيف وسلع التزلج في الشتاء...إلخ. أنظر إلى الشكل (8-1-C).

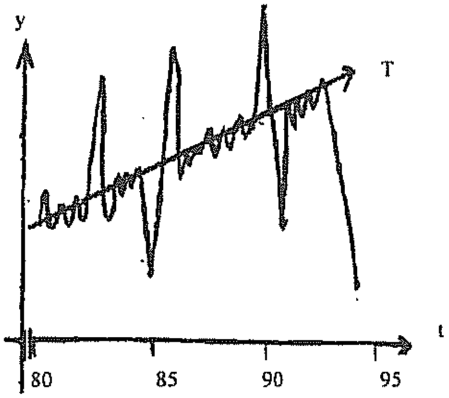
3- التغيرات الدورية (C): وتمثل كافة التبدلات والتغيرات النمطية المتكررة خلال فترات زمنية غير محددة الطول وعادة ما تكون الفترة أطول من السنة مثل دورات نمو الرخاء الاقتصادي والصناعي...إلخ. أنظر إلى الشكل (8-1-b).

4- التغيرات العشوائية (I): وتمثل التغيرات والتبدلات قيد الدراسة وغير الخاضعة لنمط معين أو فترة محددة مثل الزلازل المدمّرة أو النمو الإنتاجي بعد الحروب أو الثورات الشعبية ...إلخ . كما في الشكل (8-1-d).



الشكل (8-1-C)

يوضح التغيرات والتبدلات S خلال السنة الواحدة



الشكل (8-1-d)

يوضح التبدلات العشوائية I غير الخاضعة لفترة معينة ولا تتكرر بنفس النمط أو الأسلوب بل خاضعة للصدفة

وعادة ما تكون التغيرات في السلاسل الزمنية والناتجة عن تغيرات التوجه العام T أو التغيرات الموسمية S والدورية C هي تغيرات نمطية أي أنها لا تحدث بمحض الصدفة بل تعكس عوامل ومؤثرات ذات تأثير منتظم نوع ما بعكس التغيرات العشوائية التي تشير إلى عوامل ومؤثرات غير منتظمة أو غير متوقعة الحدوث وقد تحدث في أي فترة زمنية بمحض الصدفة وبصورة عشوائية مثل الفيضانات، الهزات الأرضية، والاضرابات العمالية. وعلى هذا الأساس تصبح دراسة السلاسل الزمنية عبارة عن تقدير أو تحديد وحساب كل من مكونات السلسلة Y الأربع ومدى تأثيرها على السلسلة (Y) خلال فترة السلسلة كلها.

8-2 نماذج السلاسل الزمنية:

لأجل فصل تأثير كل من المكونات لا بد من وضع بعض الفرضيات عن علاقة هذه المكونات ببعضها البعض، وعادة ما يطلق على مثل هذه الافتراضات أسم نموذج السلسلة الزمنية وعلى هذا الأساس تصنف السلاسل الزمنية إلى ثلاثة نماذج هي:

1- **النموذج الجمعي** $Y=T\oplus S\oplus C\oplus I$:The additive modle

2- **النموذج الضربي** $Y = T\otimes S\otimes C\otimes I$: The multiplicative modle

3- **النموذج المختلط The mixed modle**

حيث يَفترِض النموذج الجمعي بأنّ التابع أو السلسلة Y ما هو إلاّ مجموع لمكوناته الأربعة المستقلة عن بعضها البعض، بينما يفترض النموذج الضربي عدم استقلالية هذه المكونات الأربعة عن بعضها البعض. أمّا النموذج المختلط فهو قائم على افتراض استقلالية بعض المكونات وترابط بعضها الآخر مع بعضه وبهذا يأخذ هذا النموذج واحداً من الأشكال التالية:

$$Y=S\oplus T\otimes C\otimes I \quad \text{أو} \quad Y=C\oplus T\otimes S\otimes I$$

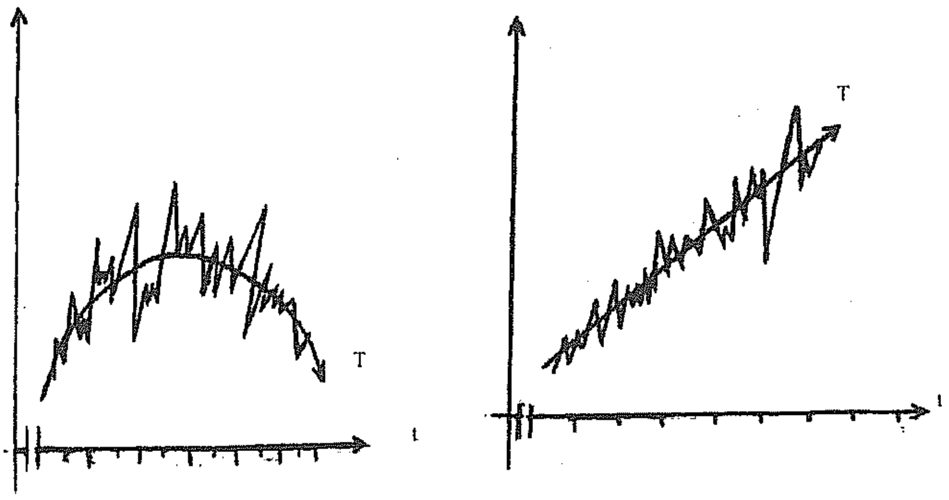
ولحساب وتحديد هذه المكونات الأربعة نعتبر أنّ كلاً من S، C، I عبارة عن مقتطفات أو أجزاء من التوجه أو النمط العام T للسلسلة حيث يتم حسابه أو تقديره أولاً ومن ثم حذفه أو اختزاله من قيمة السلسلة Y، وهذا يسمح بأن تكون المتغيرات المتبقية ناتجة عن أحد المكونات الثلاث الباقية وهي S, C, I وكل منها يمكن تقديره بوسائل إحصائية تقنية. ومتى ما تم تحديد هذه المكونات ونسبها المشتركة في السلسلة الزمنية يكون من السهل عندئذ التنبؤ بقيمة السلسلة الزمنية عند نقطة معينة أو زمن محدد في المستقبل ومن المشاكل الشائعة والهامة هو التنبؤ بحجم المكونة العشوائية I في السلسلة الزمنية والتي قد تحدث خلال فترة السلسلة مثل حدوث الأعاصير أو الثورات الشعبية أو زيادة أو انخفاض الطاقة بصورة مفاجئة. أمّا التنبؤ بالمكونات الأخرى فهو أسهل من الحالة السابقة بشرط تجانس وتوفر نفس الظروف خلال فترة السلسلة على شرط أن يكون هذا التنبؤ خلال فترة زمنية مستقبلية قصيرة.

8-3 التوجه العام (أو النمط) الخطي للسلسلة (linear Trend):

في كثير من الحالات وبعد دراسة البيانات الإحصائية يحدد منحني السلسلة الزمنية نوعية النمط الذي يكون عادة أحد نوعين: خطي أو غير خطي. ولتقدير النمط الخطي للسلسلة هنا نفترض المعادلة الخطية:

$$y = mx + b \qquad (8.1)$$

حيث x يمثل المتغير الزمني t مقاساً بوحدات معينة.



-b-
توجه أو نمط غير خطي للسلسلة الزمنية

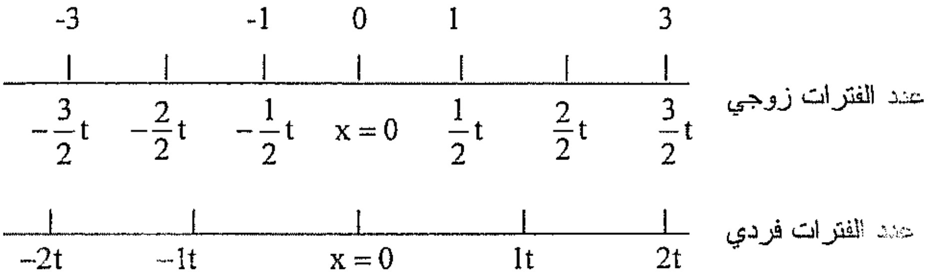
-a-
نمط خطي متصاعد للسلسلة الزمنية

الشكل (8-2)

والجدير بالملاحظة هنا هو أسلوب تدريج الاحداثي الزمني x حيث نفترض أنّ $x = 0$ هو منتصف الفترة الزمنية للقياس بالأشهر أو بالأسابيع أو السنين أياً كانت فمثلاً إذا كان القياس كل ستة اشهر فإننا نضع $x = 0$ عند منتصف الفترة بالضبط أي بعد ثلاثة أشهر فقط ومن ثم نستمر بالقياسات كل ستة أشهر يميناً ويساراً.

فإذا كان عدد الفترات فردياً فنضع $x = 0$ عند مركز الفترات مهما كانت نوعية وحدات القياس وترقم الفترات نفسها بـ $x = 0, \pm1, \pm2, \pm3 \ldots$

أمّا إذا كان عدد الفترات الزمنية زوجياً فنستخدم أنصاف الوحدات أو الفترات المستخدمة الترقيم ونضع $x = 0$ عند النقطة الوسطية بين الفترتين الوسطيتين حيث يتم بعدئذ ترقيم الفترات اليسرى بوحدات سالبة لكل نصفين بينما يتم ترقيم الفترات اليمنى بوحدات موجبة لكل نصفين أيضاً، ويمكن توضيح ذلك كما يلي:



وباستخدام طريقة المربعات الصغرى الشائعة نجد أنّ المعادلات النموذجية لهذه السلسلة تكتب بالصورة المختصرة التالية:

$$\sum_{i=1}^{n} y_i = m\sum_{i=1}^{n} x_i + n\,b \qquad \textbf{(8.2)}$$

وبضرب المعادلة (8.2) في x_i والجمع نحصل على التالي:

$$\sum_{i=1}^{n} x_i y_i = m\sum_{i=1}^{n} x_i^2 + n\,b\sum_{i=1}^{n} x_i \qquad \textbf{(8.3)}$$

وعلى أساس الافتراض السابق في التدريج نجد أنّ $\sum_{i=1}^{n} x_i = 0$ ومنه نجد أنّ:

$$\sum_{i=1}^{n} y_i - n\,b \qquad \textbf{(8.4)}$$

ومنها نجد أنّ الثابت:

$$b = \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y} \qquad \textbf{(8.5)}$$

ومن (8.3) نجد أنّ:

$$\sum_{i=1}^{n} x_i y_i = m\sum_{i=1}^{n} x_i^2$$

ومنه نجد أنّ:

$$m = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \qquad (8.6)$$

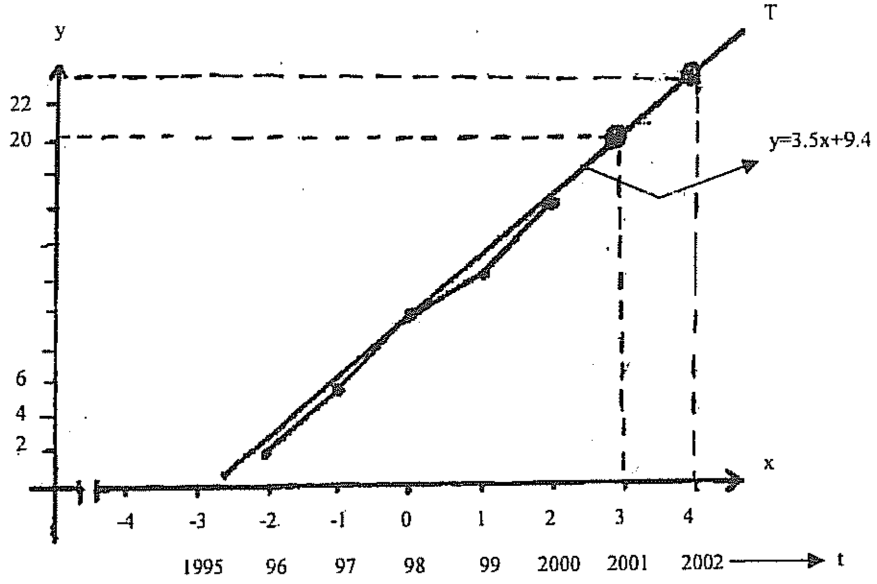
وبالتعويض في المعادلة (8.2) نحصل على أفضل نمط أو توجه عام خطي للسلسلة الزمنية.

مثال (1): إذا كانت المبيعات السنوية (y) لشركة صناعية مبينة في الجدول التالي بملايين الليرات:

السنة	1996	1997	1998	1999	2000
المبيعات	2	6	10	13	16

فارسم السلسلة الزمنية التي تمثل مبيعات الشركة وأحسب معادلة التوجه العام لهذه السلسلة.

الحل: بإتباع التوضيحات والتعليمات المختصة بالرسم نلاحظ أنّ الفترة الزمنية الكلية فردية أي خمس سنوات وعليه نبتدئ التدريج لمحور x بوضع الصفر (نقطة الأصل) عند منتصف السنة الوسطية 1998 (الأول من الشهر السابع) ومن ثم نختار الوحدات المناسبة لكل من محوري x (عند منتصفات السنين) و y كما في الشكل التالي:



الشكل 3-8: منحني السلسلة الزمنية وخط التوجه العام للسلسلة الذي يتنبأ بكمية المبيعات لعام 2002.

ولحساب المكونة الأولى T (معادلة التوجه العام للسلسلة) نحسب عناصر المعادلات النموذجية المختصرة (8.3)، (8.6) كما في الجدول التالي:

جدول (8-1) يوضح عناصر معادلتي المربعات الصغرى النموذجتين

السنوات	x_i	y_i بالملايين	x_i^2	$x_i \cdot y_i$
1996	-2	2	4	-4
1997	-1	6	1	-6
1998	0	10	0	0
1999	+1	13	1	13
2000	+2	16	4	32
	0	47	10	35

وبالتعويض في المعادلتين (8.3)، (8.4) أو مباشرة بالمعادلتين (8.5)، (8-6) نجد أنّ:

$$\sum_{i=1}^{5} y_i = nb \quad \Rightarrow \quad b = \frac{\sum y_i}{n} = \bar{y} = \frac{47}{5} = 9.4$$

وكذلك

$$\sum_{i=1}^{5} x_i y_i = m \sum_{i=1}^{5} x_i^2 \quad \Rightarrow \quad m = \frac{\sum_{i=1}^{5} x_i y_i}{\sum x_i^2} = \frac{35}{10} = 3.5$$

ومن النتائج نجد أنّ معادلة التوجه العام السنوية T تعطى بالشكل التالي:

$$y = mx + b \quad = \quad 3.5x + 9.4 \qquad \textbf{(8.7)}$$

حيث x تمثل وحدات سنوية مثبتة عند منتصف كل سنة (أي الأول من شهر تموز)

ويتضح من الشكل والمعادلة بأنّ التوجه العام للسلسلة الزمنية لمبيعات الشركة الصناعية يتمثل بخط مستقيم متصاعد ذي ميل $m = 3.5$ ويمكن استخدام معادلته الخطية للتنبؤ بما ستكون عليه كمية المبيعات في منتصف سنة 2002 مثلاً حيث نجد عندئذٍ أنّ $x = 4$ وعليه فإنّ كمية المبيعات عندئذٍ (أي في 2002/7/1) ستكون:

$y = (3.5)(4) + 9.4 = 14 + 9.4 = 23.4$ مليون ليرة

في دراستنا للسلاسل الزمنية نجد في كثير من الحالات ضرورة التحويل من الوحدات السنوية إلى الوحدات الأسبوعية أو الشهرية أو الربع سنوية...إلخ وفي مثل هذه الحالات يتوجب تقسيم كلا المعاملين a ، b وكذلك x على 52 أسبوعاً أو 12 شهراً أو 4 (ربع السنة) حسب الترتيب.

ولو طولبنا على سبيل المثال بحساب المبيعات الربع سنوية (أي كل ثلاثة أشهر) فعندئذ يتوجب تحويل معادلة التوجه العام T (8.7) لسلسلة المبيعات إلى الصورة ربع السنوية وذلك بقسمة كل من b ، m ، x على 4 لتصبح المعادلة على الصورة ربع السنوية التالية:

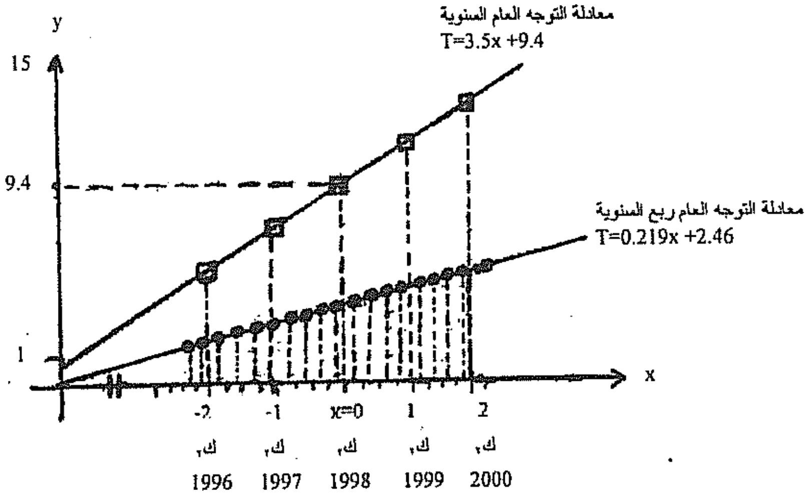
$$\hat{y}=\frac{b}{4}+\frac{m}{4}(\frac{x}{4}) \qquad (8.8)$$

حيث نستقرئ منها أنّ $x=0$ تعني الأول من تموز 1998 . بينما $x=1$ تعني الأول من تشرين الأول 1998 أي بعد ثلاثة أشهر من التاريخ السابق وهكذا كما نلاحظ بأن المتغير x والمعادلة مازالت متمركزة عند الأول من تموز وليس في وسط أحد الأرباع ولو أردنا إزاحة نقطة الوسط $x=0$ إلى منتصف الربع إلى اليمين أي بمقدار(شهر ونصف الشهر) ليصبح متمركزاً عند منتصف الشهر الثامن (15 آب) فإننا في مثل هذه الحالة نضيف $\frac{1}{2}$ إلى المتغير x ليصبح $(x+\frac{1}{2})$ بدلاً من x فقط وعليه تؤول المعادلة الخطية للتوجه العام T للصورة التالية: $\hat{y}=\frac{b}{4}+\frac{m}{16}(x+\frac{1}{2})=(\frac{b}{4})+(\frac{m}{32})+(\frac{m}{16}x)$

حيث تعني $x=1$ في هذه الحالة منتصف شهر تشرين الثاني 1998 بينما تعني $x=2$ منتصف شهر شباط أي أنّ المعادلة للمثال (1) تؤول إلى الصورة التالية:

$$\hat{y}=\frac{9.4}{4}+\frac{3.5}{32}+(\frac{3.5}{16}x)=2.46+0.219x \qquad (8.9)$$

حيث يمكن تمثيل معادلة التوجه العام الربع سنوية المتمركزة في منتصف الشهر الثامن (آب) كما في الشكل (4-8) التالي:



الشكل (8-4): حيث نلاحظ بأنّ الدوائر الصغيرة ⊗ التي تمثل قيم التوجه العام للسلسلة الزمنية ربع السنوية تتمركز في منتصفات أرباع السنين

ملاحظة ①: في بعض الأحيان تعطى قيم التوجه العام T كل ربع سنة ولكنها محسوبة بالأسلوب السنوي وفي مثل هذه الحالات نجد أنّ معادلة التوجه العام تمثل أربعة أضعاف معادلة التوجه العام الربع سنوية أي يمكن إرجاع المعادلة (8.9) أعلاه إلى الصورة السنوية كالتالي:

$$\hat{y} = 4(2.46 + 0.219(4x)) = 3.5x + 9.4$$

وهي نفس المعادلة (8.7) السابقة.

وبنفس الأسلوب يمكن معالجة تحويل قيم التوجه العام السنوية إلى شهرية وبالعكس و هذه التحولات تفرض نفسها أحياناً وفقاً لنوعية الإنتاج الذي يعتمد على موسم واحد مثلاً من كل سنة مثل صناعة السكر من قصب السكر.

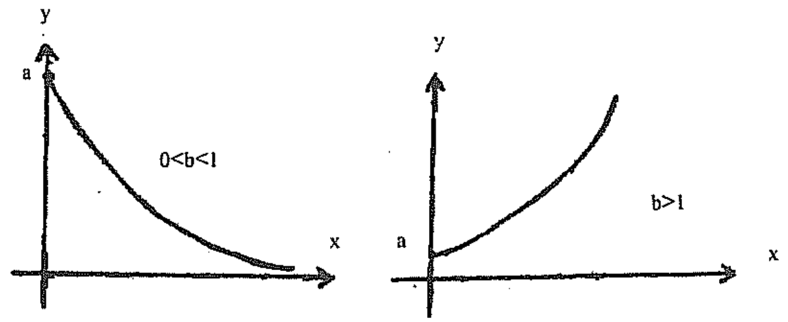
8-4 التوجه العام T (أو النمط) غير الخطي:

تختلف حالة حساب التوجه العام T غير الخطي عن الحالة السابقة إذ نحاول هنا إيجاد نموذج أو معادلة غير خطية لوصف التوجه العام في السلسلة الزمنية وهنالك العديد من النماذج غير الخطية التي يمكن استخدامها نذكر منها.

أ – نموذج منحني الدالة الأسية (Exponential Curve)

$$y = a\, b^x \qquad (8.10)$$

حيث $a, b > 0$ الذي يأخذ الصورة التالية:



الشكل (8-5)

وهذا النموذج أو المنحني يساعد في وصف السلاسل الزمنية التي تتصاعد أو تتنازل بنمط واحد بالنسبة للزمن مثل مبيعات البضائع الجديدة أو النمو السكاني وانتشار الأمراض السريعة المعدية وبأخذ لوغاريتم الطرفين (اللوغاريتم العشري) تتحول الصورة الأسية إلى معادلة خطية كالتالي:

$$\log y = \log(a\, b^x) = \log a + x \log b$$

$$= A + Bx$$

وبتطبيق طريقة المربعات الصغرى السابقة نستطيع حساب كل من $A = \log a$ $B = \log b$ لإيجاد معادلة التوجه العام الأفضل.

مثال (2): أحسب معادلة التوجه العام للسلسلة الزمنية لمبيعات شركة صناعية بملايين الليرات وفقاً للجدول التالي:

السنة	1996	1997	1998	1999	2000
المبيعات	1	3	6	14	41

الحل: بملاحظة انتشار قيم المبيعات نجد أنها تتصاعد بوتيرة عالية غير خطية وعليه نفترض النموذج الأسي الذي يمثل التوجه العام T كالتالي:

$$y = a\,b^x$$

أو

$$\log y = \log a + x \log b$$

وبحل هذه المعادلة الخطية وفقاً لطريقة المربعات الصغرى بعد اختيار تدريج x بصورة مناسبة حيث $\sum x_i = 0$ نجد من تطبيق المعادلتين (8.5) و (8.6) مباشرةً أنّ:

$$A = \log a = \frac{\sum_{i=1}^{n} \log y_i}{n} \quad , \quad B = \log b = \frac{\sum_{i=1}^{n} x_i \log y_i}{\sum_{i=1}^{n} x_i^2}$$

وبتكوين جدول العناصر التالي:

جدول (8-2)

السنة	x_i	y_i	$\log_{10} y_i$	$x_i \log_{10} y_i$	x_i^2
96	-2	1	0	0	4
97	-1	3	0.477	-0.477	1
98	0	6	0.778	0	0
99	1	14	1.146	1.146	1
2000	2	41	1.613	3.226	4
المجموع	0		4.014	3.895	10

وبالتعويض في العلاقتين أعلاه نجد أنّ:

$$A = \log_{10} a = \frac{4.014}{5} = 0.8028 \Rightarrow a = 6.35$$

$$B = \log_{10} b = \frac{3.895}{10} = 0.3895 \Rightarrow b = 2.45$$

وعليه نجد أنّ معادلة التوجه العام T تأخذ الصورة التالية كما في الشكل أدناه:

$$\hat{y} = (6.35)(2.45)^x$$

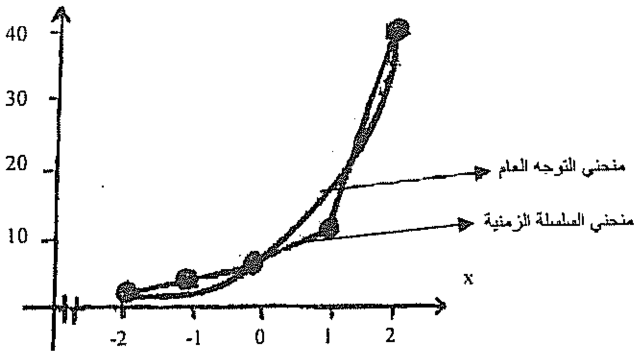
وباستخدام هذه المعادلة يمكننا التنبؤ بمقدار المبيعات لعام 2001

أي عندما $x = 3$:

$\hat{y} = (6.35)(2.45)^3 = 93.34$ (ملايين الليرات)

أو مبيعات 2002 عندما $x = 4$:

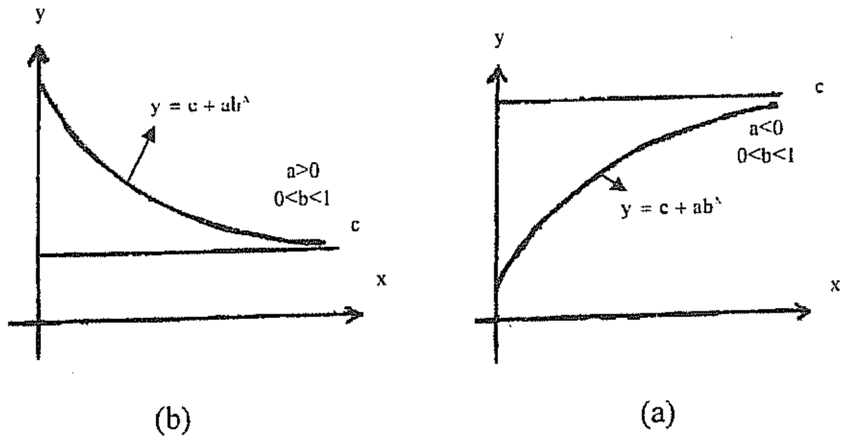
$\hat{y} = (6.35)(2.45)^4 = 228.683$ (ملايين الليرات)



الشكل (6-8)

ب- نموذج منحني الدالة الأسية المعدل: الذي يأخذ الصورة التالية، حيث يلاحظ بأنه يقارب القيمة C التي تمثل حداً أعلى أو أدنى للتوجه العام

$$\hat{y} = C + a\,b^x \qquad \textbf{(8.11)}$$

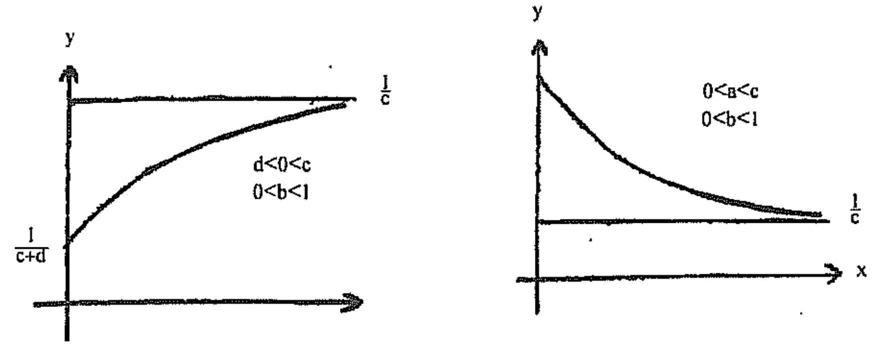


الشكل (7-8)

ملاحظة ②: في المثال (2) السابق نجد أنّ $C = 0$.

ج – **نموذج منحني التسويق logistic Curve**: والذي يأخذ الصورة التالية:

$$\hat{y} = \frac{1}{C + a\, b^{x}} \qquad \mathbf{(8.12)}$$



(b) (a)

الشكل (8-8)

ويلاحظ بأن اختيار النموذج غير الخطي لتمثيل التوجه العام لسلسلة زمنية ليس سهلاً على الرغم من ضرورته. كما يتوجب توخي الحذر في مثل هذه الحالات التي يتباعد فيها منحني نموذج التوجه العام خصوصاً عند الرغبة في التنبؤ وتقدير القيم المستقبلية حيث يكون التنبؤ غير دقيق أو مضبوط عندئذ.

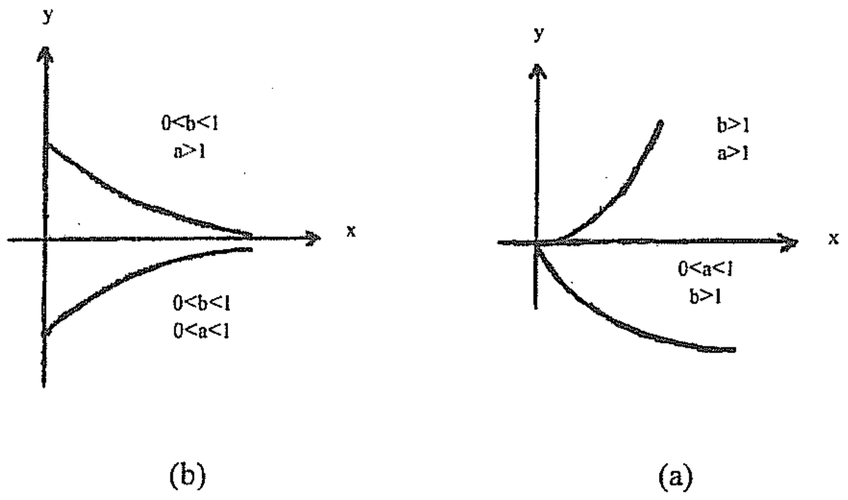
د – **منحني كمبيرتز Compertz Curve**: ويأخذ الصورة والأشكال التالية:

$$\hat{y} = c a^{b^{x}} \qquad \mathbf{(8.13)}$$

$$\log y = \log c + b^{x} \log a \qquad \mathbf{(8.14)}$$

وهو يمثل حالة خاصة من حالات المنحني الأسي حيث يوجد فيه ثلاثة وسطاء هي:

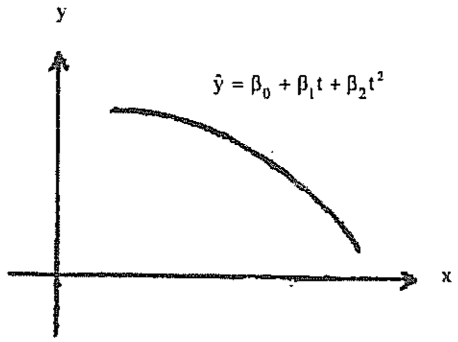
$\log a, b, \log c$



الشكل (8-9)

هـ- نموذج المقادير التربيعية: والذي يأخذ الصورة التالية:

$$\hat{y} = \beta_0 + \beta_1 t + \beta_2 t^2 \qquad \textbf{(8.15)}$$



الشكل (8-10)

إنّ استخدام أحد هذه النماذج السابقة يعتمد على قيم البيانات المستحصلة من التجارب الميدانية كما لاحظنا في المثال (2).

8-5 طريقة المتوسطات لتعديل السلسلة الزمنية:

تسمى هذه الطريقة بالمتوسطات المتحركة أيضاً وتستخدم كثيراً للتخلص من (أو تعديل) القيم الكبيرة والمتطرفة الناتجة عن المكونات الموسمية والدورية والمؤثرات اليومية أو الأسبوعية وحتى العشوائية أو الشاذة. وتعرف المتوسطات المتحركة على أنها متتالية من المتوسطات يمثل كل منها متوسط قيم السلسلة الزمنية خلال فترة (h) محددة ومتتالية على طول السلسلة الزمنية ولو حددنا الفترة الزمنية (h) بإثنا عشر شهراً (سنة) مثلاً فإنّ متوسطات الـ 12 شهراً للسلسلة الزمنية تكوّن هي الأخرى سلسلة زمنية ثانية معدّلة وخالية من القيم المتطرفة مثل القيم الموسمية التي تحدث مرة واحدة عادة كل سنة، وبالمثل نجد أنّ المتوسطات المتحركة لفترة خمس سنوات محددة ثابتة ستحذف كل القيم الشاذة (I) والمتطرفة الموسمية (S) والدورية الخمس سنوية (C).

مثال (3): أحسب السلسلة الزمنية المعدلة بالمتوسطات المتحركة لأرباح شركة صناعية بياناتها السنوية مسجلة في الجدول التالي (بملايين الليرات) وحدد معادلة التوجه العام T لها.

جدول (8-3)

السنة x	1988	89	90	91	92	93	94	95	96	97	98	99	2000	01	02
الأرباح y	2	4	5	7	8	6	8	11	13	14	11	14	18	20	23
المتوسطات M_i لكل خمس سنوات			5.2	6	6.8	8	9.2	10.4	11.4	12.6	14	15.4	17.2		

الحل: يلاحظ من البيانات وشكل انتشارها البياني بأنّ هنالك نمط دوري لتغير البيانات يتكرر خلال كل خمس سنوات على الرغم من تصاعد البيانات بشكل مطّرد. لذا نحسب المتوسطات المتحركة على أساس أن الفترة الثابتة هي خمس سنوات، وعليه نجد المتوسطات التالية:

$$M_1 = \frac{1}{5}(2+4+5+7+8) = 5.2$$

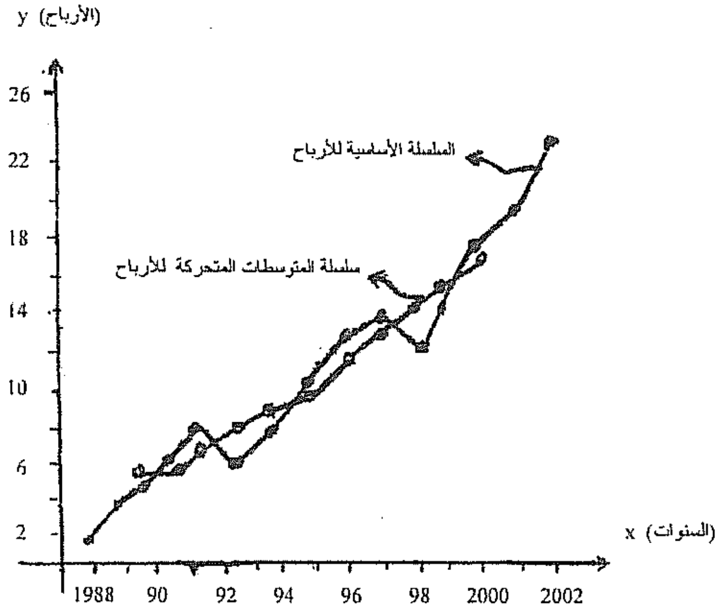
$$M_2 = \frac{1}{5}(4+5+7+8+6) = 6$$

$$M_3 = \frac{1}{5}(5+7+8+6+8) = 6.8$$

$$\vdots$$

$$M_{11} = \frac{1}{5}(11+14+18+20+23) = 17.2$$

حيث يلاحظ بأنّ هذه المتوسطات M_i تتمركز عند السنة الوسطية للفترة المحددة أي منتصف كل خمس سنوات على التتالي ويلاحظ من الشكل البياني التالي الفرق بين السلسلة الزمنية الأصلية والسلسلة الزمنية المعدلة للأرباح.



الشكل (11-8): قيم السلسلة الزمنية للأرباح الأساسية وقيم السلسلة الزمنية المعدّلة للمتوسطات المتحركة للأرباح

ملاحظة ③: يمكن استخدام الأوزان النوعية الإحصائية لإظهار أهمية ومعنوية القيمة في سنة من السنين فمثلاً لو استخدمنا الأوزان 1,2,3,2,1 على التتالي لإظهار أهمية قيمة الأرباح في السنوات الخمس المتتالية بين بعضها البعض أو بالنسبة للسوق التجارية أو الصناعية، فإنّ سلسلة المتوسطات المتحركة والموزونة تأخذ القيم التالية:

$$M_1 = \frac{1(2)+2(4)+3(5)+2(7)+1(8)}{1+2+3+2+1} = 5.22$$

$$M_2 = \frac{1(4)+2(5)+3(7)+2(8)+1(6)}{1+2+3+2+1} = 5.67$$

$$\vdots$$

$$M_{11} = \frac{1(11)+2(14)+3(18)+2(20)+1(23)}{1+2+3+2+1} = \frac{156}{9} = 17.33$$

وهذه السلسلة الجديدة تمثل تعديلاً آخر للسلسلة الأساسية للأرباح في المثـال السـابق وفقـاً للأهمية والوزن المعطى.

وعلى الرغم من أنّ سلسلة المتوسطات المتحركة في المثال (3) تعتبر خطية كما يشير لـذلك الشكل (8-11) إلاّ أنها لا تحدد معادلة التوجه العام T وعليه باسـتخدام طريقـة المربعـات الصغرى للسلسلة الزمنية المعدّلة كما في المثال (2) نحصل على معادلة التوجه العام التالية:

لحساب معادلة التوجه العام $\hat{y} = mx + a$ نكوّن الجدول التالي حيث $x = 0$ تتمركـز عنـد منتصف السنة الوسطية (1995/7/1) للسلسلة غير المعدلة.

جدول (8-4)

السنة	1988	89	90	91	92	93	94	95	96	97	98	99	2000	01	02
x_i	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
y_i	2	4	5	7	8	6	8	11	13	14	11	14	18	20	23
$x_i y_i$	-14	-24	-25	-28	-24	-12	-8	0	13	28	33	56	90	120	161
x_i^2	49	36	25	16	9	4	1	0	1	4	9	16	25	36	49

حيث نجد أنّ: $\sum x_i = 0$ و $\sum y_i = 164$ و $\sum x_i y_i = 366$ و $\sum x_i^2 = 280$

ومنه:

$$a = \frac{\sum y_i}{n} = \frac{164}{15} = 10.93 = \bar{y}$$

$$m = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{366}{280} = 1.31$$

ومنه تكون معادلة التوجه العام ا لخطية بالشكل:

$$\hat{y} = 1.31x + 10.93$$

ملاحظة ④: إنّ اختيار الفترة المحددة h لتعديل سلسلة زمنية يجب أن يتناسب مع ضـرورة التنعيم. فكلما زاد طول الفترة كلما كان التعديل أفضل ويتقارب خطياً في الشكل.

ملاحظة ⑤: لو كانت الفترة الكلية للسلسلة الزمنيـة 14 سـنة مـثلاً بـدلاً مـن 15 فـإنّ اختيار $x = 0$ سيتمركز عند منتصف الفترة الزوجية أي عند بداية عام 95(95/1/1) ونهايـة عام 94(94/12/31)

ولحساب معادلة التوجه العام المعدّلة للسلسلة المعدّلة فسوف نستخدم قيم المتوسطات المتحركة وبنفس الأسلوب كما في الجدول التالي:

جدول (8-5)

السنة	90	91	92	93	94	95	96	97	98	99	2000	المجموع
x_i	-5	-4	-3	-2	-1	0	1	2	3	4	5	$\Sigma x_i = 0$
الأرباح .	5.2	6	6.8	8	9.2	10.4	11.4	12.6	14	15.4	17.2	$\Sigma M_i = 1162$
$x_i M_i$	-26	-24	-20.4	-16	-9.2	0	11.4	25.2	42	61.6	86.0	$\Sigma x_i M_i = 1306$
x_i^2	25	16	9	4	1	0	1	4	9	16	25	$\Sigma x_i^2 = 110$

حيث يكون:

$$m = \frac{\sum x_i M_i}{\sum x_i^2} = \frac{130.6}{110} = 1.19 \qquad a = \frac{\sum M_i}{n} = \frac{116.2}{11} = 10.56 = \bar{M}$$

ومنه فإنّ معادلة التوجه العام الخطية للسلسلة المعدّلة تأخذ الصورة التالية:

$\hat{y} = 1.19x + 10.56$ وهي مقاربة للمعادلة السابقة.

ومن هاتين المعادلتين يمكن التنبؤ (بصورة عامة) بما ستكون عليه الأرباح بعد سنة أو سنتين مستقبلاً.

8-6 حساب وتقدير مركبة المؤثرات الموسمية (Seasonal Effects):

تُقدر أو تحسب المركبة الموسمية للسلسلة الزمنية باستخدام ما يسمى بقيمة النسبة المئوية الاتجاهية (ratio to-trend) والتي تمثل مقدار أو نسبة التذبذب في التغيرات الموسمية الشهرية منسوبة لقيمة التوجه العام للسلسلة أي أنّ:

القيمة النسبية المئوية الاتجاهية **(8-16)** $$S_{Rt} = \frac{y}{\hat{y}} \times 100$$

حيث y تمثل القيمة الأساسية للسلسلة الزمنية عند الزمن t_0

$\hat{y}$ تمثل القيمة المحسوبة الناتجة عن التوجه العام T عند الزمن t_0

وعليه فإنّ القيمة 105٪ في السلسلة تعني زيادة 5٪ عن التوجه العام وبالعكس نجد أنّ القيمة 92٪ في السلسلة تعني نقصان بمقدار 8٪ عن قيمة التوجه العام.

مثـــال (4): أحسب التأثيرات الموسمية للسلسلة الزمنية التي تمثل المبيعات الشـــهرية بـــآلاف الليرات لشركة صناعية كما في الجدول التالي:

الجدول (8-6)

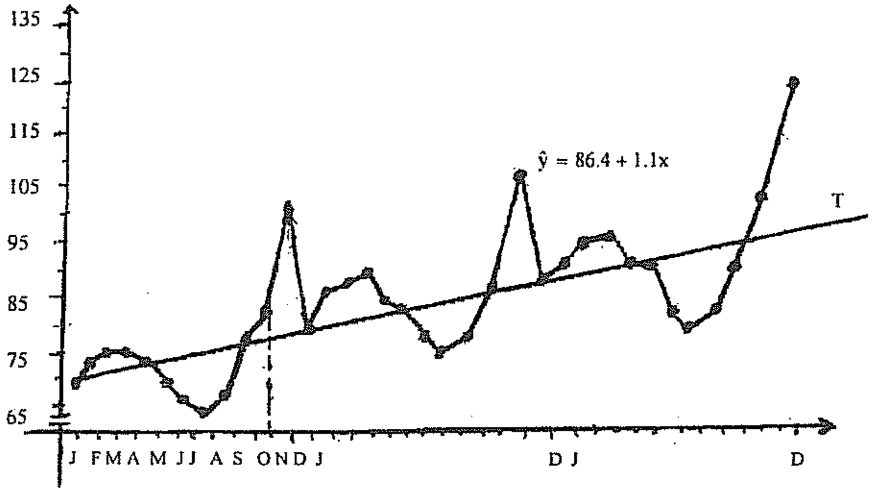
الأشهر / السنوات	ك٢ 1	شباط 2	آذار 3	نيسان 4	أيار 5	حزيران 6	تموز 7	آب 8	أيلول 9	ت١ 10	ت٢ 11	ك١ 12	متوسط المبيعات الشهرية $\overline{Y}$
2000	67	71	72	73	71	70	67	64	66	74	82	99	73.0
2001	80	86	89	89	84	83	79	76	77	87	98	114	96.8
2002	92	98	101	103	96	96	89	87	90	99	112	131	99.5

الحل: نلاحظ من انتشار المبيعات الشهرية (شكل (8-12)) بأنّ التوجه العـــام لسلســـلة القـــيم الشهرية تأخذ الطابع أو النمط الخطي وعليه نحسب المعادلة الخطية السنوية للتوجه العـــام T باستخدام المربعات الصغرى (أو أي طريقة أخرى للانحدار) كالتالي:

السنوات	x_i	$\overline{y}_i$	x_i^2	$x_i\overline{y}_i$
2000	-1	73.0	1	-73
2001	0	96.8	0	0
2002	+1	99.5	1	99.5
المجموع	0	259.3	2	26.5

ومنه نجد أنّ:

$$m=\frac{\sum x_i\overline{y}_i}{\sum x_i^2}=\frac{26.5}{2}=13.3 \quad , \qquad b=\frac{\sum \overline{y}_i}{n}=\frac{259.3}{3}=86.44$$



الشكل (8-12)

ومنها نجد أنّ معادلة التوجه العام السنوية هي:

$$\hat{y} = 13.3x + 86.44 \qquad (8.17)$$

وللتعبير عن المتغير x بالأشهر يكفي التقسيم على 12 لتصبح المعادلة كالتالي:

$$\hat{y} = 13.3\left(\frac{x}{12}\right) + 86.44 = 1.108x + 86.44 \qquad (8.18)$$

$$= 1.1x + 86.44$$

القيم المحسوبة من العلاقة الشهرية المعدلة (8.18) تتمركز في بداية كل شهر لأنّ $x = 0$ تتمركز في بداية الشهر السابع من العام 2001(2001/7/1) ولأجل الحصول على القيمة المتوسطة في منتصف الشهر السابع من العام نفسه مثلاً نعوض $x = \frac{1}{2}$ لنحصل على القيمة التالية:

$$\hat{y} = 86.4 + 1.108\left(\frac{1}{2}\right) = 87.0$$

وعلى أساس هذه القيمة يمكننا تقدير كافة القيم لمتوسطات الأشهر السابقة واللاحقة لعام 2001 وبالمثل للأعوام 2000، 2002 كما في الجدول التالي وذلك بطرح 1.1 شهرياً للأشهر السابقة وإضافة 1.1 شهرياً للأشهر اللاحقة على الترتيب:

جدول (8-7)

(جدول التوجه العام T الشهري للمبيعات)

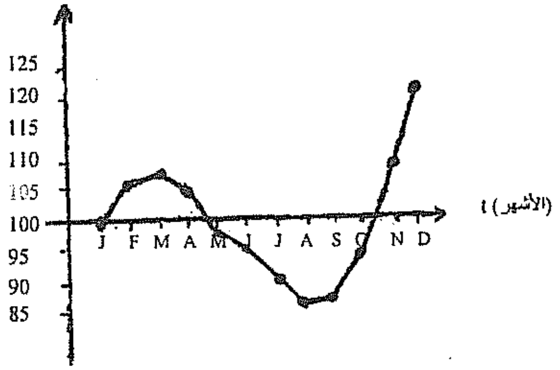
الأشهر / السنوات	ك2/15 1	شباط/15 2	آذار/15 3	نيسان/15 4	أيار/15 5	حزيران/15 6	تموز/15 7	آب/15 8	أيلول/15 9	ت1/15 10	ت2/15 11	ك1/15 12
2000	67.2	68.3	69.4	70.5	71.6	72.7	73.8	74.9	76.0	77.1	78.2	79.3
2001	80.4	81.5	82.6	83.7	84.8	85.9	87.0	88.1	89.2	90.3	91.4	92.5
2002	43.6	94.7	95.8	96.9	98.0	99.1	100.2	101.3	102.4	103.5	104.6	105.7

ولحساب النسبة المئوية لتأثير التبدلات الموسومية على التوجه العــام T للمبيعــات الشــهرية نستخدم العلاقة (8.17) أعلاه وذلك بقسمة كل قيمة في الجدول (8-6) على المناظرة لها فــي الجدول (8-7) وذلك للتخلص من تأثيرات التوجه العام T: $S_{Rt} = \frac{y}{\hat{y}} \times 100$

لنحصل على الجدول (8-8) التالي الذي يمثل نسب التأثيرات الموسمية في المبيعات الشهرية:

جدول (8-8)

الأشهر / السنوات	ك2 1	شباط 2	آذار 3	نيسان 4	أيار 5	حزيران 6	تموز 7	آب 8	أيلول 9	ت1 10	ت2 11	ك1 12
2000	99 7	103.9	103.7	103.5	99.1	96.2	90.7	85.4	86.8	95.9	104.8	124.8
2001	99.5	105.5	107.7	106.3	99.0	96.6	90.8	86.2	86.3	96.3	107.2	123.2
2002	98.2	103.4	105.4	106.2	97.9	96.8	88.8	85.8	87.8	95.6	107	123 9



الشكل (8-13): المنحني البياني للنسب المئوية الموسمية التي تبين متوسط التأثيرات أو التبدلات الموســمية فــي المبيعات الشهرية للسنوات الثلاث.

حيث تمثل كل قيمة من قيم الجدول السابق مقدار الارتفاع أو الانخفاض (الزيادة أو النقصان) على النسبة 100٪ التي تمثل التوجه العام T للمبيعات ونلاحظ بأنّ هنالك ثلاث قيم نسبية للتأثيرات الموسمية لكل شهر وعليه لتوحيد هذه النتيجة نحسب المتوسط الشهري لهذه القيم النسبية الثلاث حيث نحصل على القيم التالية التي تمثل التبدلات أو التأثيرات الموسمية بصورة أفضل وكما يلي:

جدول (9-8)

متوسط التأثيرات الموسمية على المبيعات الشهرية

الأشهر	ك٢ 1	شباط 2	آذار 3	نيسان 4	أيار 5	حزيران 6	تموز 7	آب 8	أيلول 9	ت١ 10	ت٢ 11	ك١ 12
متوسطات الأشهر	99.1	104.3	105.6	105.3	98.7	96.5	90.1	85.8	87.0	95.9	106.3	124

ويلاحظ من الجدول السابق بأن التأثيرات الموسمية لشهر كانون الثاني أدنى من المستوى العام 100٪ بمقدار 0.9٪ بينما تزداد هذه التأثيرات بمقدار 4.3٪ لشهر شباط (فبراير) وبمقدار 5.6٪ لشهر آذار بينما تزداد أو أعلى بمقدار 24٪ لشهر كانون الأول.

ملاحظة ⑥: يمكن استخدام جدول المتوسطات المتحركة أو جدول المتوسطات المركزية المتحركة لحساب وتقدير التأثيرات الموسمية حيث نلاحظ انعدام التأثيرات أو التغيرات العشوائية بشكل كبير في هذه الطريقة. حيث تصبح القاعدة الحسابية بدلاً من (8.16) كالتالي: قيمة التأثيرات الموسمية منسوبة للمتوسطات المتحركة:

$$S_{RM} = \frac{y}{\bar{y}} \times 100 \qquad \textbf{(8.19)}$$

حيث $\bar{y}$ يمثل المتوسطات المتحركة للسلسلة عند الزمن t_0 بينما y يمثل القيمة الفعلية للسلسلة الزمنية عند الزمن t_0. وبالنسبة لهذه الطريقة فإنها تتلخص بأخذ المتوسط الأثنا عشري الشهري للسلسلة الزمنية والذي يتم فيه انعدام التأثيرات العشوائية I حيث يصبح النموذج الضربي خالياً من التأثيرات العشوائية أي أنّ: $Y = T \times S \times C$ فقط وبقسمة القيمة الأصلية في السلسلة على $T \times C$ نحصل على التأثيرات الموسمية S. أي أنّ:

$$\frac{Y}{T \times C} = \frac{T \times S \times C}{T \times C} = S$$

وباستخدام نفس السلسلة الزمنية في المثال (4) السابق نجد أنّ المتوسطات المتحركة والمتوسطات المركزية تأخذ القيم المجدولة تالياً.

جدول (10-8)

السنة	الأشهر	المبيعات بآلاف الليرات	المتوسطات المتحركة الاثنا عشرية	المتوسطات المركزية الاثنا عشرية
2000	ك2 1	67		
	شباط 2	71		
	آذار 3	72		
	نيسان 4	73		
	آيار 5	71		
	حزيران 6	70	73.0	
	تموز 7	61	74.1	73.6
	آب 8	64	75.3	74.7
	أيلول 9	66	76.8	76.1
	ت1 10	74	78.1	77.5
	ت2 11	82	79.2	78.7
	ك1 12	99	80.3	79.8
2001	ك2 1	80	81.3	80.8
	شباط 2	86	82.3	81.8
	آذار 3	89	83.3	82.8
	نيسان 4	89	84.3	83.8
	آيار 5	84	85.7	85.0
	حزيران 6	83	86.9	86.3
	تموز 7	79	87.9	87.4
	آب 8	76	88.9	88.4
	أيلول 9	77	89.9	89.4
	ت1 10	87	91.2	90.6
	ت2 11	98	92.1	91.7
	ك1 12	114	93.2	92.7
2002	ك2 1	92	94.0	93.6
	شباط 2	98	94.9	94.5
	آذار 3	101	96.0	95.5
	نيسان 4	103	97.0	96.5
	آيار 5	96	98.2	97.6
	حزيران 6	96	99.6	98.9
	تموز 7	89		
	آب 8	87		
	أيلول 9	90		
	ت1 10	99		
	ت2 11	112		
	ك1 12	131		

الجدول (8-11)

جدول التقديرات الموسمية ومتوسطاتها

الأشهر / السنوات	ك٢ 1	شباط 2	آذار 3	نيسان 4	أيار 5	حزيران 6	تموز 7	آب 8	أيلول 9	ت١ 10	ت٢ 11	ك١ 12
2000	-	-	-	-	-	-	91.0	85.7	86.9	95.5	104.1	124.0
2001	99.0	105.1	107.1	106.1	98.8	96 2	90.4	86.0	86.2	96.0	106.7	122.9
2002	98.4	103.9	105.9	106.8	98.5	97.0	-	-	-	-	-	-
المتوسطات	98.7	104.5	106.5	106.5	98 7	96.6	90.7	85.9	86.6	95.8	105.4	123.5

8-7 حساب وتقدير مركبة المؤثرات الدورية
:(Determining The Cyclical Effects)

لحساب المركبة الدورية (C) آخذين بعين الاعتبار النمــوذج الضـــربي $Y = T \times S \times C \times I$ حيث يتم حذف المكونة S باستخدام أسلوب القسمة على النسبة للمتوسطات المتحركة، حيــث تقسم قيمة $\hat{y}$ عندئذٍ على النسبة المئوية الموسمية $\left(\frac{S}{100}\right)$ وحيث نحصل عندئــذٍ علــى قــيم النموذج $T \times C \times I = \frac{\hat{y} \times 100}{S}$ وباستخدام طريقة المتوسطات الصغرى (الصورة الخطية أو غير الخطية) لتحديد قيم المكونة T ومن ثم قسمة قيم النموذج مرة أخرى على قيم T المناظرة نحصل على قيم النموذج $\hat{y} = C \times I$ ولفصل قيم المكونة العشوائية I عن القــيم والمــؤثرات الدورية C فإننا نختار أوزان مناسبة فردية العدد (ثلاثية أو خماسية الفترات) ويفضل أن يكون مجموعها مساوياً للواحد ولتكن مثلاً 0.1- ,0.3 ,0.6 ,0.3 ,0.1- متمركــزة حــول وســطها والفائدة من هذا التوزين هو الحصول على تمثيل أفضل لمنحني المكونة الدورية C. فمثلاً لو أردنا تقدير أو حساب قيمة المكونة C في شهر أبريل (نيسان) فنجد أنه يساوي التالي:

$$C_{April} = C_4 = -0.1(C \times I)_{Feb} + 0.3(C \times I)_{March} + 0.6(C \times I)_{April} + 0.3(C \times I)_{May} - 0.1(C \times I)_{June} \quad (8.20)$$

وبنفس الأسلوب نحصل على تقديرات المكونة C للأشهر الأخرى حيث تتمركز قيمة نيســان وسط الأوزان الخمسة المذكورة أعلاه.

ففي الجدول (8-10) تم حذف واختزال المؤثرات الموسمية S مع بعض المؤثرات العشوائية I من السلسلة وللحصول على تقدير المؤثرات الموسمية، نقسم القيم الأساسية في السلسلة علــى

قيم المتوسطات المتحركة النصف شهرية المناظرة لها لنحصل عندئذٍ على التقديرات الموسمية كما هي مدرجة في الجدول (8-11). وبقسمة القيم الأساسية (y) لمبيعات الشركة في السلسلة على التقديرات الموسمية المناظرة لها (عناصر الجدول (8-11)) نحصـــل علـــى التقـــديرات المتبقية $\frac{y \times 100}{S} = T \times C \times I$ كما في الجدول التالي:

جدول (8-12): جدول تقديرات $T \times C \times I$ وحساب التوجه العام بالمربعات الصغرى

السنة	الأشهر	المبيعات عند منتصف كل شهر $T \times C \times I = \frac{y \times 100}{5}$	المبيعات عند y_i بداية كل شهر	x_i	x_i^2	$x_i y_i$	تقديرات T الجديدة	$c \times I = y_i / T$	C
2000	ك، 1								
	شباط 2								
	آذار 3								
	نيسان 4								
	أيار 5								
	حزيران 6								
	تموز 7	74.0							
	آب 8	74.7	74.35	-11	121	-817.85	75.88	0.980	
	أيلول 9	75.95	75.33	-10	100	-753.3	76.85	0.980	
	ت، 10	77.49	76.72	-9	81	-690.48	77.82	0.986	0.986
	ت، 11	78.77	78.13	-8	64	-625.04	78.80	0.992	0.992
	ك، 12	79.84	79.31	-7	49	-555.17	79.77	0.994	0.994
2001	ك، 1	80.81	80.33	-6	36	-481.98	80.75	0.995	0.995
	شباط 2	81.83	81.32	-5	25	-406.60	81.72	0.995	0.939
	آذار 3	83.1	82.47	-4	16	-329.88	82.69	0.997	0.997
	نيسان 4	83.88	83.49	-3	9	-250.47	83.67	0.998	0.998
	أيار 5	85.02	84.45	-2	4	-168.90	84.64	0.998	0.998
	حزيران 6	86.28	85.65	-1	1	-85.65	85.62	1.000	1.000
	تموز 7	87.39	86.84	0	0	0	86.59	1.002	1.002
	آب 8	88.37	87.88	+1	1	87.88	87.56	1.004	1.004
	أيلول 9	89.33	88.85	+2	4	177.70	88.54	1.004	1.004
	ت، 10	90.63	89.98	+3	9	269.94	89.51	1.005	1.005
	ت، 11	91.85	91.24	+4	16	364.96	90.49	1.008	1.008
	ك، 12	92.76	92.31	+5	25	461.55	91.46	1.009	1.009
2002	ك، 1	93.5	93.13	+6	36	549.78	92.43	1.008	1.008
	شباط 2	94.32	93.91	+7	49	657.37	93.41	1.005	1.005
	آذار 3	95.37	94.85	+8	64	758.80	94.38	1.005	1.005
	نيسان 4	96.44	95.91	+9	81	863.19	95.36	1.006	1.006
	أيار 5	97.46	96.95	+10	100	969.5	96.33	1.006	
	حزيران 6	98.97	98.22	+11	121	1080.42	97.30	1.009	
	تموز 7								
	آب 8								
	أيلول 9								
	ت، 10								
	ت، 11								
	ك، 12								
			$\Sigma y_i = 1991.62$	$\Sigma x_i = 0$	$\Sigma x_i^2 = 1012$	$\Sigma x_i y_i = 985.77$			

ومن أعمدة الجدول السابق (12-8) الأخيرة نستطيع حساب معادلة خط التوجه العام T الجديدة حيث نجد أنّ الثابت:

$$b = \frac{\sum y_i}{n} = \frac{1991.62}{23} = 86.59 \quad \text{والميل هو} \quad m = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{985.77}{1012} = 0.974$$

ومن النتائج أعلاه نجد أنّ المعادلة هي:

$$T : \hat{y} = mx + b = 0.974x + 86.59 \qquad \textbf{(8.21)}$$

وبحساب قيم التوجه العام من المعادلة الأخيرة نحصل على العمود الثامن [جدول (12-8)] وبقسمة عناصر العمود الرابع (متوسط المبيعات y_i) على العناصر المناظرة لها من العمود الثامن (تقديرات التوجه العام T) نحصل على تقدير للمؤثرات الدورية $C \times I$ متضمنة بعضاً من التأثيرات العشوائية I (العمود التاسع)

وباستخدام التوزين الثلاثي أو الخماسي المناسب مثل (0.1-, 0.3, 0.6, 0.3, -0.1-) (0.1, 0.2, 0.4, 0.2, 0.1) , (0.3, 0.4, 0.3) لحذف التأثيرات العشوائية المتبقية نجد وفقاً للتوزين الخماسي الأول بأنّ: تقدير التأثير الدوري على المبيعات (ومن ضمنها على وحدات السلع) للشهر العاشر (تشرين الأول) مثلاً هو:

$$C_{10} = -0.1(C\times I)_8 + 0.3(C\times I)_9 + 0.6(C\times I)_{10}$$
$$+0.3(C\times I)_{11} - 0.1(C\times I)_{12}$$
$$= -0.1(0.98) + 0.3(0.98) + 0.6(0.986) + 0.3(0.992) - 0.1(0.994)$$
$$= -0.098 + 0.294 + 0.592 + 0.298 - 0.099 = 0.987$$

ومن النتيجة الأخيرة يمكن استقراء التأثيرات الدورية على المبيعات أو الوحدات السلعية لشهر (10) أو تشرين أول.

وبنفس الأسلوب يمكن تقدير قيم التأثيرات الدورية للأشهر المتبقية كما هو مدرج في العمود العاشر من الجدول (12-8). ويلاحظ اختفاء التقديرات الدورية للشهرين الثامن والتاسع من السلسلة في بداية العمود والشهرين الخامس والسادس للسلسلة في نهاية العمود العاشر نتيجة لعملية التوزين الخماسي الذي أُعتمد في الحسابات. ولو أردنا الحصول على تلك التقديرات الدورية للأشهر المحذوفة لتوجب عندئذٍ زيادة فترة أو طول السلسلة الزمنية أي زيادة عدد البيانات لتشمل سنوات أخرى قبل سنة 2000 وبعد سنة 2002 مثلاً كما يلاحظ عينات التأثيرات العشوائية I نظراً لتساوي القيم المتناظرة في العمودين 9 ، 10 من الجدول (12-8).

8-8 الأرقام القياسية (Index numbers):

تعتبر الأرقام القياسية وسيلة من وسائل المقارنة العددية الهامة بين كميتين مثل α،β من صنف واحد أي ضمن مجموعة واحدة تسمى إحداها الكمية أو العدد الأساسي أي الكمية التي تم اعتبارها أساساً للمقارنة أو المقايسة. وعادةً ما تكون الكميات الأساس ذات معنى أو أهمية متعارفة أو مرجعية تاريخية غير منكرة...إلخ مثل سنوات الجفاف أو سنوات المعارك أو سنوات الرخاء والفيضانات...إلخ وعلى ضوء ما تقدم وعلى اعتبار أن الكمية β هي كمية الأساس فإنّ الرقم القياسي للكمية α والذي يرمز له عادة بالرمز I_α منسوباً للكمية الأساس β يحسب ويعرف بالعلاقة المئوية التالية:

$$I_\alpha = \frac{\alpha}{\beta_{Base}} \times 100 \qquad (8.22)$$

أي أنّ الرقم القياسي I لكمية معينة مثل α يعرف بالنسبة المئوية لها مع الأساس وبلغة الاستقلالية يُعتبر α مثلاً لمتغير غير مستقل لمزاوجته أو مقارنته بالمتغير المستقل أو الثابت (الأساس) β.

وعلى أساس التعريف السابق نجد أهمية الأرقام القياسية وسعة مجالات استخدامها حيث يستخدم كمختصر مفيد لتوضيح تحسن أو عدم تحسن خطة زمنية معينة أو لتوضيح زيادة أو نقصان الإنتاجية لمصنع معين أو لدول معينة أو لشركة ما. فمثلاً الرقم القياسي للناتج القومي NPI يعرف ويحسب من العلاقة التالية:

$$I_{NP} = \frac{\text{كمية الإنتاج القومي للفترة المطلوبة (بعد الحرب مثلاً)}}{\text{كمية الإنتاج القومي لفترة أساس (قبل الحرب مثلاً)}} \times 100 \qquad (8.23)$$

ومن خلال هذه النسبة المئوية يمكننا معرفة ما إذا كان الإنتاج القومي مرتفعاً أم منخفضاً أو ثابتاً وهو ما يعكس تحسن الأداء والإدارة والمردود النقدي والعمالة أو البطالة ...إلخ

ومن العلاقة (8.22) السابقة يمكننا معرفة الكمية الحقيقية المرجعية الأساس β حيث:

$$\beta = \frac{\alpha}{I_\alpha} \times 100 \qquad (8.24)$$

وهذه علاقة مفيدة للمقارنة الحقيقية بين المتغيرات المستقلة المختلفة فمثلاً لو ارتفع مرتب موظف من 10 000 ليرة شهرياً إلى 15 000 ليرة شهرياً في الوقت الذي ارتفع فيه الرقم القياسي للدخل من 100٪ إلى 130٪ خلال السنوات 2000 إلى 2003 فإننا نلاحظ بأنّ:

المرتب أو الدخل الحقيقي لعام 2000 هو: $\beta_1 = \frac{10000}{100} \times 100 = 10000$

المرتب أو الدخل الحقيقي لعام 2003 هو: $\beta_2 = \frac{15000}{130} \times 100 = 11,538$

أي أنّ الزيادة الحقيقية في المرتبات هي 1538 ليرة وليس 5000 ليرة كما هو متبادر للذهن في الوهلة الأولى وذلك لعدم ثبوت الأساس. وعلى هذا يتوجب ثبوت الأساسات كي تتم المقارنة الحقيقية.

ولو قيل لنا أنّ الرقم القياسي الحالي لمنتج سلعي هو 95٪ عن الرقم الأساسي وهذا يعني أنّ هنالك نقصاً بمقدار 5٪ عن كمية الإنتاج في السنة الأساس أو في المصنع الأساس ولو افترضنا أنّ كمية الإنتاج في السنة الأساس كانت 950 وحدة فإنّ مقدار النقص هو:

$$950 \times \frac{5}{100} = 47.5 \approx 48$$

وحدة منتجة. وبالمثل لو كان الرقم القياسي هو 115٪ فإنّ ذلك يعني زيادة في كمية الإنتاج عن قرينتها في السنة أو المصنع الأساس بمقدار 15٪ أي ما يعادل ($146 \approx 145.5$ وحدة منتجة) ومن الجدير بالملاحظة أنّ هذا النقصان قد يعكس خصائص سلبية أخرى مثل التلكؤ الإداري، نقص المواد الأولية، نقص الطاقة والعمالة ...إلخ وبالعكس نجد أنّ الزيادة قد تعكس خصائص ايجابية مثل توفر المواد الأولية وزيادة الطلب وحسن التسويق وكذلك الجودة ...إلخ.

وكمثال ثالث على دراسة وأهمية الأرقام القياسية: نفترض أنّ الدولة خصصت مبلغ 135 مليون ليرة لإقامة وبناء مجمع صناعي في ميزانية العام 1995 حيث كانت الأسعار والأجور وسعر الأرض والمعدات ...إلخ قد تم حسابها وفقاً لرقم قياسي للتكلفة الإعمارية هو 125٪ نسبة إلى أسعار التكلفة في فترة أو سنة الأساس لعام 1985 حيث الرقم القياسي 100٪: ولو افترضنا أنّ الدولة قد خصصت ميزانية 145 مليون ليرة لإقامة مشروع صناعي آخر في العام 2003 (حيث الرقم القياسي لتكلفة الإعمار قد ارتفع إلى 180٪ نسبة إلى نفس سنة الأساس السابقة 1985) فإنّ القيمة الحقيقية لهذه الميزانيات المخصصة تكون كالتالي:

القيمة الحقيقية للمخصصات عام 1995 بملايين الليرات هو:

$$\frac{135}{125} \times 100 = 108 \text{ مليون ليرة}$$

القيمة الحقيقية للمخصصات عام 2003 بملايين الليرات هو:

$$\frac{145}{180}\times 100 = 80.56 \text{ مليون ليرة}$$

وهذا يعني أنّ المخصصات للإعمار في ميزانية عام 2003 وهي 145 مليون لا تكفي إلاّ لبناء وإعمار 75٪ تقريباً من المشروع.

وعلى هذا الأساس نلاحظ ضرورة إدخال الأرقام القياسية مهما كانت قيمها في حساب الميزانيات المخصصة للمشاريع القصيرة والطويلة المدى على السواء حيث نجد أنّ:

المبالغ المخصصة للمشروع عام 2003 يتوجب أن تكون كالتالي:

$$\frac{135\times 180}{100} = 243 \text{ مليون ليرة}$$

بدلاً من 145 مليون ليرة وذلك نظراً للتضخم الموجود في أسعار التكلفة أو النمو الاقتصادي...إلخ

ومن تعريف الرقم القياسي نستخلص الخواص التالية:

1- نسبة الرقم القياسي: $R_{\alpha/\beta} = \frac{\alpha}{\beta_{Basc}} = I_\alpha \times \frac{1}{100}$

-2 $R_{\alpha/\alpha} = 1$

-3 $R_{\alpha/\beta} \times R_{\beta/\alpha} = 1 \rightarrow R_{\alpha/\beta} = 1/R_{\beta/\alpha}$

-4 $R_{\alpha/\beta} \times R_{\beta/\gamma} = R_{\alpha/\gamma}$

ويُستفاد من هذه الخاصية الأخيرة في تعديل أو تبديل الأرقام القياسية من سنة أساس β مثلاً إلى سنة أساس أخرى γ وفقاً للعلاقة التالية للنسبة المئوية بين الرقمين القياسيين:

-5 $I_{\alpha/\gamma} = \frac{I_{\alpha/\beta}}{I_{\gamma/\beta}}\times 100 = \frac{R_{\alpha/\beta}}{R_{\gamma/\beta}}\times 100 = R_{\alpha/\gamma}\times 100$

حيث γ هي الأساس الجديد.

8-8-1 الأرقام القياسية للأسعار (price index numbers):

تلعب الأرقام القياسية للأسعار دوراً هاماً في الحسابات العملية للدول أو الشركات والمصانع وحتى الفرد في المجتمع لما لها من انعكاسات اقتصادية هامة فزيادة الرقم القياسي للأسعار لبضائع معينة ومختلفة قد يعني التضخم الاقتصادي وبعكسه قد يعني الركود والجمود بكافة النواحي الاقتصادية والإنتاجية ذات العلاقة.

ويعرّف الرقم القياسي للسعر على أنه النسبة المئوية لسعر سلعة معينة خلال زمن معين إلى سعر نفس السلعة أو المادة المنتجة التي تعتبر السلعة الأساس خلال زمن أساس أي أنّ:

$$I_p = \frac{P_i}{p_{Base}} \times 100 \qquad (8.25)$$

وعليه نجد أنّ الرقم القياسي للسعر الأساس في السنة الأساس يساوي $I_{Base} = 100\ \%$ دائماً

وعلى أثر الأهمية والحاجة لاستخدام الأرقام القياسية في النواحي الاقتصادية والإنتاجية تعددت أنواع هذه الأرقام حيث نذكر منها:

1- **الرقم القياسي التجميعي البسيط** **Simple aggregate** I_p

2- **الرقم القياسي المتوسط البسيط** **Simple average** I_p

3- **الرقم القياسي المتوسط الموزون** **Weighted average** I_p

وعادة ما تستخدم هذه الأرقام لمقارنة مجموعة مختلفة من أسعار السلع خلال سنة معينة مع أسعار نفس المجموعة من السلع خلال السنة الأساس وهو ما يسمى أحياناً بسعر السوق السلعية.

ويعرف الرقم القياسي التجميعي البسيط بالنسبة المئوية لمجموع أسعار منتجات مختلفة إلى مجموع أسعارها خلال السنة الأساس فلو افترضنا أنّ الأسعار الحالية لمجموعة n من السلع خلال عام 2004 هي: $P_1 P_2 ... P_n$ وأنّ أسعار نفس المجوعة من السلع خلال السنة الأساس 2000 مثلاً هي $P_{01} P_{02} ... P_{0n}$ فإنّ الرقم القياسي التجميعي البسيط I_n للأسعار يُحسب وفقاً للعلاقة التالية:

$$I_{np} = \frac{\sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} P_{0i}} \times 100 \qquad (8.26)$$

أمّا الرقم القياسي المتوسط البسيط فيعرّف بمتوسط الأرقام القياسية لأسعار مجموعة مؤلفة من n من السلع المختلفة ويحسب بالعلاقة التالية:

$$\overline{I}_p = \frac{\sum_i I_{Pi}}{n} = \frac{\sum_i \left(P_i / P_{0i} \right)}{n} \qquad (8.27)$$

أمّا الأرقام القياسية المتوسطة الموزونة فهي تأخذ بنظر الاعتبار أهمية السلعة في السوق التجارية حيث تقاس هذه الأهمية بمقدار الأموال المصروفة على هذه السلعة خلال السنة وهو

يساوي $P_i\,q_i$ حيث تمثل q_i كمية الإنتاج أو عدد السلع المستهلكة خلال تلك السنة وعلى ضوء عملية التوزين هذه تعددت أيضاً الأرقام القياسية المتوسطة الموزونة فمنهم من يستخدم الأهمية الأساس $(P_0\,q_0)$ ومنهم من يستخدم الأهمية للسلع خلال سنة القياس وليس الأساس.

ويُعرف الرقم القياسي المتوسط الموزون بأهمية السلع الأساس (أي بجداء الأسعار والكميات للسنة الأساس) بالعلاقة التالية:

$$I_{nB\omega}=\frac{\sum_{i=1}^{n}\left(\frac{P_i}{P_{0i}}\right).\left(P_{0i}\,q_{0i}\right)/N}{\sum_{i=1}^{n}P_{0i}\,q_{0i}/N}\times 100 \qquad \textbf{(8.28)}$$

أي النسبة المئوية لمتوسط مجموع النسب الموزونة بالأهمية الأساس $P_{0i}\,q_{0i}$ علـــى متوســـط مجموع الأوزان الأساس. أمّا الرقم القياسي المتوسط الموزون بأهمية السلع غير الأساس (أي بجداء الأسعار والكميات الحالية $P_i\,q_i$ وليس الأساسية) فيُعرّف بالعلاقة التالية:

$$I_{n\omega}=\frac{\sum_{i=1}^{n}\left(\frac{P_i}{P_{0i}}\right)P_i\,q_i}{\sum_{i=1}^{n}P_i\,q_i}\times 100 \qquad \textbf{(8.29)}$$

أي النسبة المئوية لمتوسط مجموع النسب الموزونة بالأهمية غير الأساس $P_i\,q_i$ على متوســـط مجموع الأهميات (الأوزان $P_i\,q_i$) أي (مختصراً) متوسط مجموع الأرقام القياســـية الموزنـــة علي متوسط مجموع الأوزان. ولتوضيح العلاقات السابقة نستعرض الأمثلة التالية:

مثـــال (5): أحسب الأرقام القياسية لإنتاج تلفزيونات سيرونكس خلال الأعوام المجدولة تاليـــاً معتبراً إنتاج سنة 1980 هو الأساس:

السنوات	1970	1975	1980	1985	1990	2000
كمية الانتاج بالاف الوحدات	70	79	90	91	75	112

الحل:

$$I_{70}=\frac{70}{90}\times 100=77.78\,\%$$

$$I_{75}=\frac{79}{90}\times 100=87.78\,\%$$

$$I_{80}=\frac{90}{90}\times 100=100\,\%$$

$$I_{85} = \frac{91}{90} \times 100 = 101.11\,\%$$

$$I_{90} = \frac{75}{90} \times 100 = 83.33\,\%$$

$$I_{2000} = \frac{112}{90} \times 100 = 124.44\,\%$$

ويلاحظ من النتائج بأن هنالك نقصاناً على المستوى الإنتاجي المطلوب بمقـدار ‎22٪‎ و ‎12٪‎ و ‎17٪‎ تقريباً خلال الأعوام 70 و 75 و 90 على الترتيب وهو ما يعكس سلبية معينـة فـي الإنتاج قد تكون إدارية أو تلكؤ في الطاقة أو العمالة أو نقص المواد الأوليـة ...إلـخ. بينمـا تعكس الأرقام الأخرى عكس ذلك بالنسبة لسنة الأساس 1980.

مثال (6): أحسب كلاً من الرقم القياسي التجميعي البسيط والرقم القياسي المتوسط البسيط لسنة 2001 لمجموعة من السلع معتبراً 1987 سنة أساس ومستخدماً البيانات المجدولة تالياً:

جدول (13-8)

أنواع المنتجات	الوحدات q_0	P_{0i} **1987**	P_i **2001**
الحليب	2 kg	25	35
الجبنة	1 kg	60	80
الزبدة	½ kg	30	45
لبن مصفى	1 kg	35	55
القشطة	1 kg	180	250
المجموع	5½ kg		

الحل: نوجد أولاً الأرقام القياسية الأحادية، ثم نوجد مجموع الأسعار الأساس ومجموع الأسعار لعام 2001 بوحداتها المتعارفة:

(للحليب) $I_p = \frac{17.5}{12.5} \times 100 = 140\,\%$

(الجبنة) $I_p = \frac{80}{60} \times 100 = 133.33\,\%$

(الزبدة) $I_p = \frac{90}{60} \times 100 = 150\,\%$

ويلاحظ أيضاً من الشكل بأنّ العملية الإنتاجية منضبطة أو محكمة.

9-7 خرائط ضبط المدى والمتوسط المشتركة ($\bar{X}$ and R charts):

يمكننا الاستفادة من حسابات خرائط الجودة والمراقبة لكل من المدى R والمتوسط $\bar{X}$ لتكوين خرائط مشتركة بينها. فلو استخدمنا $\bar{\bar{x}}$ كتقدير لمتوسط المجتمع μ واستخدمنا $\hat{\sigma}=\frac{\bar{R}}{d_2}$ كتقدير غير منحاز لانحراف المجتمع المعياري σ فإنّ حدود خرائط المدى والمتوسط المشتركة تأخذ القيم التالية (حسب الاختيار المتعارف 3σ):

$$UCL=\bar{\bar{x}}+3\frac{\hat{\sigma}}{\sqrt{n}}=\bar{\bar{x}}+\frac{3}{\sqrt{n}}\frac{\bar{R}}{d_2}=\bar{\bar{x}}+A_2\bar{R} \qquad \textbf{(9.10)}$$

الحد الأوسط (خط الاعتدال)
$$CL=\bar{\bar{x}}$$

$$LCL=\bar{\bar{x}}-3\frac{\hat{\sigma}}{\sqrt{n}}=\bar{\bar{x}}-\frac{3}{\sqrt{n}}\frac{\bar{R}}{d_2}=\bar{\bar{x}}-A_2\bar{R} \qquad \textbf{(9.11)}$$

حيث: $A_2=\frac{3}{d_2\sqrt{n}}$ وهي قيم محسوبة في جداول خاصة أيضاً.

وأنّ: $\hat{\sigma}=\frac{\bar{R}}{d_2}=\frac{0.023}{2.326}=0.0099$

مثال (3): استخدم البيانات في المثال (1) لتكوين خارطة المدى والمتوسط المشتركة.

الحل: عرفنا من المثالين (1) و (2) بأنّ: $\bar{\bar{x}}=74.001=CL \quad , \quad \bar{R}=0.023$

ولحساب حدود الضبط العليا والدنيا نجد من الجداول بأنّ:

$$A_2=\frac{3}{d_2\sqrt{n}}=\frac{3}{\sqrt{5}\,d_2}=0.577$$

$$UCL_{\bar{x},R}=\bar{\bar{x}}+A_2\bar{R}=74.001+(0.577)(0.023)=74.014$$

$$LCL_{\bar{x},R}=\bar{\bar{x}}-A_2\bar{R}=74.001-(0.577)(0.023)=73.988$$

9-8 خرائط ضبط الانحراف المعياري والمتوسط المشتركة ($\overline{X}$ and S charts) :

عندما تكون أحجام العينات كبيرة ولنقل مثلاً $n > 10$ نجد أنّ عملية تقدير σ باستخدام المدى R تفقد كفاءتها نظراً لضياع الكثير من البيانات بين حدي المدى الواسع R وعلى هذا الأساس يفضل في مثل هذه الحالة استبدال خرائط $\overline{X}$ و R بخرائط $\overline{X}$ و S.

وعلى افتراض أنّ تباين العملية الإنتاجية σ^2 غير معلوم لذا نستخدم التقدير المستقل وغير المنحاز للعينة.

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} \qquad \textbf{(9.12)}$$

ولكن S ليس تقديراً غير منحاز للانحراف المعياري σ وعلى هذا الأساس نجد أنّ S عبارة عن تقدير لوسيط مثل $c_4\sigma$ أي أنّ $E(s) = c_4\sigma$ حيث c_4 معامل التقدير:

$$c_4 = \left(\frac{2}{n-1}\right)^{\frac{1}{2}} \frac{\Gamma(n/2)}{\Gamma(\frac{n-1}{2})} \qquad \textbf{(9.13)}$$

وهو ثابت محسوب ومجدول يعتمد على حجم العينة العشوائية n أمّا تباين الانحرافات المعيارية S للعينات فيقدر أو يحسب من العلاقة التالية $\sigma\sqrt{1-c_4^2}$. وهذه المعلومة تستخدم في حساب حدود خرائط ضبط S.

وعلى افتراض أنّ σ معلومة فإنّ حدود الضبط لخرائط S تكتب كالتالي:

$$CL = c_4\sigma \qquad \textbf{(9.14)}$$

$$UCL = c_4\sigma + 3\sigma\sqrt{1-c_4^2} = B_6\sigma \qquad \textbf{(9.15)}$$

حيث: $B_6 = c_4 \text{-} 3\sqrt{1-c_4^2}$ قيم محسوبة ومجدولة مسبقاً.

$$LCL = c_4\sigma - 3\sigma\sqrt{1-c_4^2} = B_5\sigma \qquad \textbf{(9.16)}$$

حيث: $B_5 = c_4 \text{-} 3\sqrt{1-c_4^2}$ قيم مجدولة مسبقاً.

وعند عدم معلومية σ يتم أخذ m من العينات العشوائية كل منها ذات حجم n لتقدير σ وذلك باستخدام متوسط الانحرافات $\overline{S}$ حيث نجد منها أنّ:

$$\overline{S} = \frac{1}{m}\sum_{i=1}^{m} s_i \qquad \textbf{(9.17)}$$

حيث s_i يمثل الانحراف المعياري للعينة رقم i.

ومنها نجد أنّ الإحصائية $\overline{S}/c_4$ تمثل تقديراً غير منحاز للانحراف σ وعليه فإنّ حدود الضبط والمراقبة لخرائط S تحسب بالعلاقات التالية:

$$CL = \overline{S} = c_4\sigma$$

$$UCL = \overline{S} + 3\frac{\overline{S}}{c_4}\sqrt{1-c_4^2} = \overline{S}(1+\frac{3}{c_4}\sqrt{1-c_4^2}) = B_4\overline{S} \qquad (9.18)$$

$$LCL = \overline{S} - 3\frac{\overline{S}}{c_4}\sqrt{1-c_4^2} = \overline{S}(1-\frac{3}{c_4}\sqrt{1-c_4^2} = B_3\overline{S} \qquad (9.19)$$

وعند استخدام $E(\frac{\overline{S}}{c_4}) = \sigma$ فإنّ حدود الضبط في خرائط $\overline{x}$ تعرف وتحدد بالتالي:

$$UCL = \overline{\overline{x}} + \frac{3\overline{S}}{c_4\sqrt{n}} = \overline{\overline{x}} + A_3\overline{S} \qquad (9.20)$$

$$CL = \overline{\overline{x}}$$

حيث $A_3 = 3/c_4\sqrt{n}$ قيم محسوبة ومجدولة.

$$LCL = \overline{\overline{x}} - \frac{3\overline{S}}{c_4\sqrt{n}} = \overline{\overline{x}} - A_3\overline{S} \qquad (9.21)$$

وعلى ضوء ما سبق يمكننا تكوين خرائط ضبط التباين S^2 والتي تعتمد مباشرة على حساب تباين العينات كالتالي:

$$CL = \overline{S}^2$$

$$UCL = \frac{\overline{S}^2}{n-1}\chi^2_{(\alpha/2,n-1)} \qquad (9.22)$$

$$LCL = \frac{\overline{S}^2}{n-1}\chi^2_{(1-\alpha/2,n-1)} \qquad (9.23)$$

9-9 خرائط ضبط النسبة P (P charts):

تعتبر خرائط ضبط النسبة من مجموعة خرائط ضبط المواصفات غير الكمية مثل تالف أو غير تالف، يعمل أولاً يعمل، معيب أو غير معيب، ويُقصد بالنسبة هنا هو عدد المنتجات المعطوبة في العينة إلى عدد المنتجات فيها مثلاً أو عدد المنتجات غير المطابقة للمواصفات إلى عدد المنتجات الكلي في العينة وتستخدم هذه الخرائط عادةً عندما يراد ضبط ومراقبة

أعداد التالف أو المعطوب في العملية الإنتاجية ويتم تكوين هذه الخرائط بنفس الأسلوب السابق ويُراعى فيها عدد العينات $m \geq 25$ وبحجم $n \geq 50$ لضمان تحقق تقارب توزيع النسب من التوزيع الطبيعي.

ومن التعريف السابق نلاحظ بأنّ خرائط النسبة تعتمد بطبيعتها على ما يسمى بالتوزيع الثنائي (Binomial distribution). فلو افترضنا أنّ نسبة التالف أو المعطوب أو غير المطابق للمواصفات الفنية هي $\overline{p}$ وأنّ الوحدات المنتجة مستقلة وغير مترابطة ضمن عملية الإنتاج نجد أنّ كل وحدة من وحدات الإنتاج تمثل قيمة عشوائية لمتغير برنولي ثنائي التوزيع وبالنسبة (أو الوسيط) p. فإذا تم أخذ عينة عشوائية ذات حجم n من الوحدات وبفرض d هو عدد الوحدات غير المطابقة للمواصفات فإنّ D يعتبر متغيراً عشوائياً ثنائي التوزيع $B(n,p)$ ذو وسيطين هما (n,p) وعليه فإنّ:

$$P[D=x]=\binom{n}{x}p^{x}(1-p)^{n-x} \qquad x=0,1,\ldots,n \qquad \mathbf{(9.24)}$$

ومن معلوماتنا السابقة حول التوزيع الثنائي وتوزيع برنولي نجد أنّ المتوسط $\mu = np$ وأنّ التباين $S^2 = npq = np(1-p)$. ولو افترضنا أنّ نسبة التالف أو المعطوب في كل عينة هو $\hat{p} = \frac{d}{n}$ فإنّ المتوسط هو $\mu = p$ والتباين هو $\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$

وعلى ضوء معطيات خرائط ضبط شيوارت السابقة يمكننا صياغة حدود الضبط العليا والدنيا وخط الوسط مفترضين معلومية p مسبقاً كالتالي:

$$UCL = p + 3\sqrt{\frac{p(1-p)}{n}} \qquad \mathbf{(9.25)}$$

$$CL = p$$

$$LCL = p - 3\sqrt{\frac{p(1-p)}{n}} \qquad \mathbf{(9.26)}$$

حيث اخترنا حدود الضبط ثلاث وحدات انحرافية معيارية كالمتعارف. وعند عدم معرفة نسبة المعطوب أو التالف (p) عموماً فيمكن تقديرها وحسابها من العينات المأخوذة عشوائياً بعدد $m \geq 25$ على الأقل وكل منها بحجم $n \geq 50$ وبديهي أن يؤدي زيادة حجم العينة إلى دقة في نسبه المعطوب أو التالف من المنتجات في العينة الواحدة. ولو افترضنا في هذه الحالة أن النسبة لكل عينة هي:

$$\hat{p}_i = \frac{d_i}{n}$$

حيث d_i عدد المعطوب أو التالف في العينة رقم i $(i=1,2,\ldots,m)$
وعليه يمكن حساب متوسط النسب في هذه العينات:

$$\bar{p}=\frac{\sum_{i=1}^{m} d_i}{mn}=\frac{\sum_{i=1}^{m}\hat{p}_i}{m} \qquad (9.27)$$

حيث المتوسط أو الإحصائية $\bar{p}$ تعتبر تقديراً غير منحازاً للنسبة p غير المعلومة للمجتمع الإحصائي أي: $\mu=E(\bar{p})=p$ وكذلك $\sigma_{\hat{p}}=\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ يعتبر تقديراً للانحراف المعياري للمجتمع أي: $\sigma=E(\sigma_{\hat{p}})$ وعليه يمكننا تكوين خارطة ضبط النسبة p وحساب حدود الضبط العليا والدنيا والوسط (مفترضين عدم معلوميه p) كالتالي:

$$UCL_p=\bar{p}+3\sigma_{\hat{p}}=\bar{p}+3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \qquad (9.28)$$

$$CL_p=\bar{p}$$

$$LCL_p=\bar{p}-3\sigma_{\hat{p}}=\bar{p}-3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \qquad (9.29)$$

مثال (4): في مراقبة ميدانية لمكائن تعبئة الحليب في علب ورقية سميكة لأحدى الشركات الصناعية حيث كان المراد مراقبة وضبط تسرب الحليب من جوانب أو أسفل هذه العلب وتكوين خارطة نسبة المعيب من هذه العلب (أي العلب القابلة للتسريب)، تم أخذ 30 عينة عشوائية من هذه العلب المملوءة كل منها يحتوي على 50 علبة حليب مملوءة على فترات كل نصف ساعة وعلى كافة مراحل عمل المصنع الثلاث. ورصدت النسب كما يلي:

جدول (3-9)

رقم العينة	عدد العلب القابلة للتسريب d_i (المعيبة)	نسبة المعيب في العينة $\hat{p}_i = \frac{d_i}{n}$
1	12	0.24
2	15	0.30
3	8	0.16
4	10	0.20
5	4	0.08
6	7	0.14
7	16	0.32
8	4	0.18
9	14	0.28
10	10	0.20
11	5	0.10
12	6	0.12
13	17	0.34
14	12	0.24
15	22	0.44
16	8	0.16
17	10	0.20
18	5	0.10
19	13	0.26
20	11	0.22
21	20	0.40
22	18	0.36
23	24	0.48
24	15	0.30
25	9	0.18
26	12	0.24
27	7	0.14
28	13	0.26
29	9	0.18
30	6	0.12
	المجوع = 347	$\overline{p} = 0.2313$

الحل: من الجدول المبين أعلاه نلاحظ أنّ:

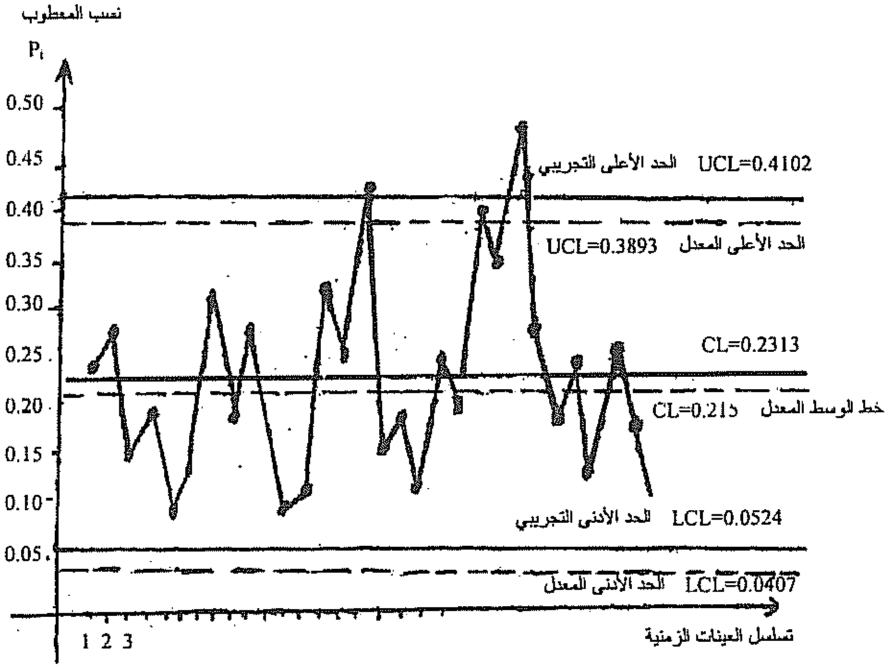
$$\overline{p}=\frac{\sum_{i=1}^{30}\hat{p}_i}{30}=0.2313=CL$$

وبالاستفادة من العلاقات (9.29)،(9.28) نجد أنّ:

$$UCL_p=\overline{p}+3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}=0.2313+0.1789=0.4102$$

$$LCL_p=\overline{p}-3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}=0.2313-0.1789=0.0524$$

وبهذه المعلومات وجدول البيانات السابق يمكننا تكوين خارطة الضبط لنسب الإنتاج المعيب كالتالي:



الشكل (9-14): يوضح خارطة ضبط نسب المعطوب من علب التعبئة لكل عينة من العينات العشوائية المـــأخوذة من المصنع وفقاً لجدول البيانات السابق.

يلاحظ من خارطة الضبط السابقة، شكل (9-14)، بأنّ هنالك نسبتين خارجتان عن حد الضبط الأعلى التجريبي وعليه نستنتج بأنّ العملية الإنتاجية ليست ضمن السيطرة أو غير محكمة ويتوجب دراسة هاتين العينتين (العينة 15 والعينة 23) لتحديد أسباب هذا الخلل واتخاذ الإجراء المناسب. وعلاوة على ذلك نجد أنّ انتشار أو تشتت نسب المعطوب عن خط الوسط CL كبيراً على الرغم من أنّ العملية الإنتاجية ضمن حدود الضبط تقريباً وهو ما ينذر بخطورة تسريب العلب للحليب وبدراسة كل من العينتين 15، 23 وجد أنّ سبب الخروج الأول عن الانضباط هو تغيير نوعية العلب خلال النصف ساعة الخامسة عشرة بينما وجد أنّ سبب الخروج الثاني عن الانضباط هو تعيين عامل جديد لآلة التعليب.

وعلى ضوء هذه الأسباب غير الميكانيكية نجد عدم ضرورة عمل صيانة للآلة وأنّ الحل يكمن فقط في حذف أو إهمال العينتين 15 ، 23 من خارطة الضبط حيث أنّ إهمالهما وغض النظر عنهما هو الحل الصائب لخارطة الضبط السابقة وعلى ضوء الإجراء المذكور يتوجب رسم حدود الضبط الجديدة بدون العينتين 15 ، 23 وعليه نجد أنّ خطوط الضبط المعدّلة (الجديدة) (revised control lines) لنسب المعطوب في المثال السابق تأخذ القيم التالية:

خط الوسط المعدّل:
$$\bar{p} = \frac{301}{(28)(50)} = 0.2150$$

الحد الأعلى المعدّل:
$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.2150 + 3\sqrt{\frac{0.2150(0.7850)}{50}} = 0.3893$$

الحد الأدنى المعدّل:
$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.2150 - 3\sqrt{\frac{0.2150(0.7850)}{50}} = 0.0407$$

وقد تمّ رسمها بشكل خطوط نقطية في نفس خارطة الضبط لنفس المثال (الشكل (9-14)).

ويلاحظ من الشكل الجديد أيضاً بأنّ نسبة المعطوب في العينة رقم 21 أصبح خارج خط الانضباط العلوي المنقط وبدراسة البيانات مرة أخرى والعينة أيضاً لم نجد سبباً منطقياً أو مقبولاً وغير ميكانيكي لهذا الخروج عن خط الضبط العلوي وبالتالي لا يمكننا إهمال أو غض النظر عن هذه النسبة الخارجة عن الانضباط وعلى ضوء ذلك وبالإضافة إلى نوعية التشتت للنسب $\bar{p}$ حول خط الوسط المعدّل CL نستنتج ضرورة صيانة آلة التعليب قيد الدراسة.

لاحظنا أنّ سبب ا لخروج الأول عن خط الانضباط العلوي لنسب العينتين 15 ، 23 كان إدارياً وليس ميكانيكياً وهو نوعية الورق وتعيين عامل جديد بدون خبرة بينما سبب الخروج الثاني عن خط الانضباط العلوي كان ميكانيكياً وليس إدارياً وبعد التعديل الإداري والصيانة الميكانيكية تم أخذ 24 عينة عشوائية حجم كل منها (50 منتجاً) وعلى فترات زمنية متساوية ($\frac{1}{2}$ ساعة خلال الدوريات الثلاث للآلة). وتم تسجيل أعداد المعيب في كل عينة مع حساب نسبها في الجدول (4-9) التالي:

جدول (4-9)

المجموعة الثانية للعينات: نسب المعطوب في العينات المأخوذة بعد التعديل والصيانة

النسب للمنتجات المعطوبة $\hat{p}_i$	عدد المنتجات المعطوبة d_i	تسلسل العينات i
0.18	9	1
0.12	6	2
0.24	12	3
0.10	5	4
0.12	6	5
0.08	4	6
0.10	5	7
0.06	3	8
0.14	7	9
0.12	6	10
0.04	2	11
0.08	4	12
0.06	3	13
0.12	6	14
0.10	5	15
0.08	4	16
0.16	8	17
0.10	5	18
0.12	6	19
0.14	7	20
0.10	5	21
0.12	6	22
0.06	3	23
0.08	4	24
المتوسط $\overline{p}_2 = 0.1108$	المجوع = 133	

ويمكننا هنا من الناحية النظرية وقبل تكوين خارطة أو مخطط الضبط لمجموعة العينات الثانية (بعد التعديل والصيانة) إجراء اختبار الفرضيات لمجموعتي العينات قبل وبعد التعديل والصيانة وهما:

1- فرضية العدم (أي عدم الاختلاف بينهما) $H_0 : \bar{p}_1 = \bar{p}_2$

2- فرضية البديل (الاختلاف بين المجموعتين) $H_1 : \bar{p}_1 > \bar{p}_2$

(أو أنّ متوسط النسب لأحدهما أكبر من الثاني على الأقل)

حيث وجدنا أنّ متوسط النسبة في مجموعة العينات الأولى قبل التعديل والصيانة هو $\bar{p}_1 = 0.215$ بينما نجد أنّ متوسط النسب بعد التعديل والصيانة في مجموعة العينات الثانية هو:

$$\bar{p}_2 = \frac{\sum_{i=1}^{m} d_i}{(50)(24)} = \frac{133}{1200} = 0.1108$$

أمّا الدرجة المعيارية للاختبار [راجع اختبار الفرضيات (الفصل السادس)]

$$z_0 = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

حيث تكون النسبة التجميعية للمنتجات المعطوبة:

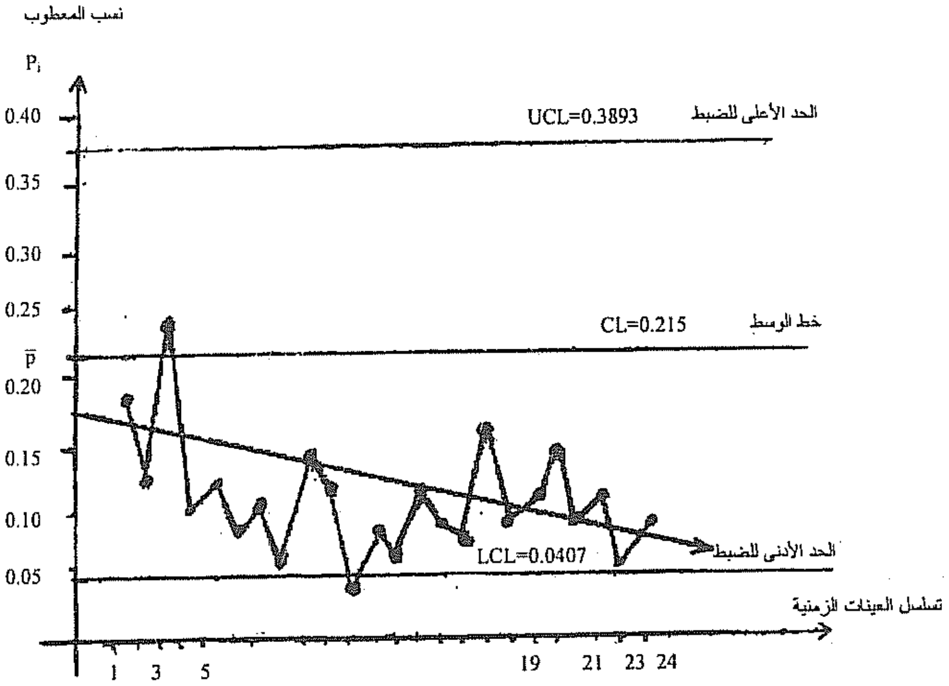
$$\hat{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{(1400)(0.215) + (1200)(0.1108)}{1400 + 1200} = 0.1669$$

$$z_0 = \frac{0.215 - 0.1108}{\sqrt{(0.1669)(0.8331)(\frac{1}{1400} + \frac{1}{1200})}} = 7.1078$$

وبالمقارنة مع الدرجة المعيارية الحرجة z_α لمنطقة الرفض $\alpha = 0.05$ المتعارفة نجد أنّ:

$$z_0 = 7.1078 > z_{\alpha = 0.05} = 1.645$$

وعليه نرفض فرض التساوي H_0 ونقبل الفرض البديل H_1 لأنّ z_0 تقع داخل منطقة الرفض أي نستنتج بأنّ هنالك نقصاناً جوهرياً لا يمكن غض النظر عنه في نسب منتجات المجموعة الثانية كما يوضحه مخطط الضبط للمجموعة الثانية (الشكل (9-15)) والمحسوبة بنفس الأسلوب السابق للمجموعة الأولى تالياً:



الشكل (9-15): مخطط الضبط لنسب المعطوب p_i في مجموعة العينات الثانية بعد الصيانة

ويلاحظ من مخطط الضبط أعلاه، شكل (9-15)، بأنّ هنالك انحداراً حاداً في خط التوجه العام لسلسلة النسب أي أنّ هنالك نقصاناً متسارعاً في نسب المنتجات المعطوبة p_i لكل عينة عن خط الوسط والاعتدال مما ينذر بخطر خروج العملية عن خط الضبط الأدنى علماً بأنّ نسبة المعطوب (التالف) في العينة 11 من المجموعة الثانية هي خارج حد الضبط الأدنى. ويبدو من خارطتي الضبط للمجموعتين الأولى والثانية، الشكل (9-14) والشكل (9-15) ضرورة إعادة النظر بحدود الضبط وأخذ مجموعة ثالثة من العينات تتكون من 40 عينة متتالية كل نصف ساعة وعلى مدى الورديات الثلاث ويحتوي كل منها على 50 عنصراً حيث تمّ تسجيل كل من أعداد المنتجات غير المتطابقة مع المواصفات (المعطوبة) ونسبها في الجدول (9-5) التالي:

جدول (5-9) المجموعة الثالثة من العينات للمثال (4) بعد التعديل والصيانة

تسلسل العينات i	أعداد المعطوب d_i	نسب المنتجات المعطوبة $\hat{p}_i = \frac{d_i}{n}$
1	8	0.16
2	7	0.14
3	5	0.10
4	6	0.12
5	4	0.08
6	5	0.10
7	2	0.04
8	3	0.06
9	4	0.08
10	7	0.14
11	6	0.12
12	5	0.10
13	5	0.10
14	3	0.06
15	7	0.14
16	9	0.18
17	6	0.12
18	10	0.20
19	4	0.08
20	3	0.06
21	5	0.10
22	8	0.16
23	11	0.22
24	9	0.18
25	7	0.14
26	3	0.06
27	5	0.10
28	2	0.04
29	1	0.02
30	4	0.08
31	5	0.10
32	3	0.06
33	7	0.14
34	6	0.12
35	4	0.08
36	4	0.08
37	6	0.12
38	8	0.16
39	5	0.10
40	6	0.16
	المجوع = 218	المتوسط $\overline{p}_3 = 0.1108$

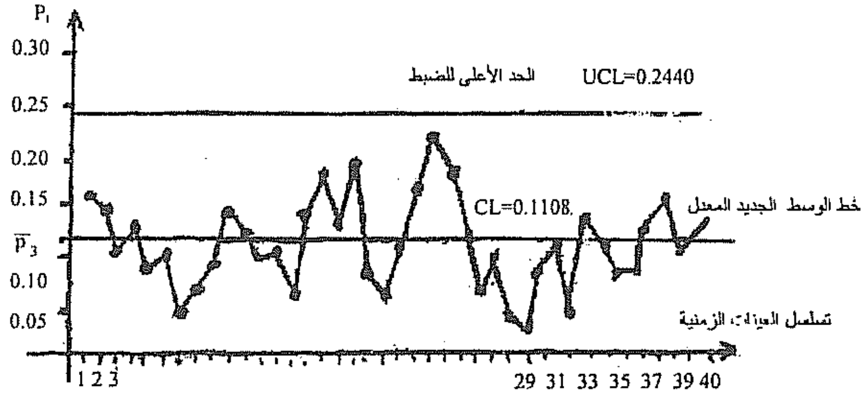
ومنه نجد أنّ حدود الضبط الجديدة المعدّلة هي كالتالي وكما في مخطط الضبط المبيّن بالشكل (9-16):

$$CL = \overline{p}_3 = 0.1108 = \overline{p}_3$$

$$UCL = \overline{p}_3 + 3\sqrt{\frac{\overline{p}_3(1-\overline{p}_3)}{n}} = 0.1108 + 3\sqrt{\frac{(0.1108)(0.8892)}{50}} = 0.2440$$

$$LCL = \overline{p}_3 - 3\sqrt{\frac{\overline{p}_3(1-\overline{p}_3)}{n}} = 0.1108 - 3\sqrt{\frac{(0.1108)(0.8892)}{50}} = -0.0224$$

ونظراً لكون خط الضبط الأدنى أصغر من الصفر $LCL < 0$ لذا نهمل أو نغض النظر عن القيمة السالبة ونستعيض عنها بخط الضبط الصفري أي اعتماد $LCL = 0$. وبهذا يكون لخريطة أو مخطط الضبط الجديد للعملية الإنتاجية حداً علوياً فقط وكما في المخطط المبيّن في الشكل (9-16) التالي:



الشكل (9-16): مخطط (خارطة) الضبط للمجموعة الثالثة من العينات في المثال رقم (4) لنسب المعطوب من علب التعبئة.

ويلاحظ من المخطط (9-16) بأنّ العملية الإنتاجية منضبطة (محكمة) للمرحلة الحالية كما نلاحظ أنّ نسب المعطوب للعينات تتقارب مع خط الوسط أو خط الاعتدال على الرغم من أنّ نسبة التسريب (أو التالف أو المعطوب من العلب) عالية أيضاً – بخلاف المطلوب وهو عدم التسريب مطلقاً – ولكنها تعتبر مقبولة.

إنّ خرائط نسب المعطوب $p_i = \frac{d_i}{n}$ من الإنتاج في العينة الواحدة تعتمد على ثلاثة عناصر:

1- حجم العينة 2- زمن وترتيب أخذ العينة 3- بعد حدّي الضبط الأعلى والأدنى عن بعضهما أو لنقل تشتت النسب عن خط الوسط أو خط الاعتدال المطلوب وفي حالة خرائط النسب $p\,(p-\text{charts})$ يتوجب أن يكون حجم العينة كبيراً لضمان الحصول على احتمال أكبر في اكتشاف وقوع منتجات معطوبة في العينة وبما أننا نرغب في عدم الحصول على حدٍ أدنى سالب لضبط الجودة أي نرغب في أن يكون:

$$LCL = p - k\sqrt{\frac{p(1-p)}{n}} > 0 \qquad \textbf{(9.30)}$$

لذا نجد أنّ:

$$n > \frac{1-p}{p}k^2 \qquad \textbf{(9.31)}$$

حيث k يمثل عدد وحدات الانحراف المعياري σ للمجتمع الإحصائي ذي التوزيع الثنائي ولو افترضنا أنّ $p = 0.05$ مثلاً، وأنّ حد الضبط المستخدم هو $k\sigma = 3\sigma$ حسب المتعارف فإنّ حجم العينة في هذه الحالة يتوجب أن يكون:

$$n > \frac{0.95}{0.05}(3^2) = 171$$

ومن هنا نلاحظ بأنّ نظرية الحد الأدنى الموجب تفيد في حساب قانون حجم العينة في خرائط النسب.

ملاحظة ①: بما أننا نتعامل مع نسب المنتجات المعطوبة في العينة الواحدة $p_i = \frac{d_i}{n}$ يمكننا استخدام أحجام مختلفة (n_i) للعينات بدلاً من الحجم الواحد الثابت n أي أنّ p_i تكون بالشكل $p_i = \frac{d_i}{n_i}$ وفي مثل هذه الحالة المتغيرة الأحجام n_i سنحصل على حدود متغيرة أيضاً لضبط الجودة والمراقبة لكل عينة من العينات وبهذا سنحصل على حد أعلى وأدنى للضبط متدرج أو متموج وغير مستقيم حيث نجد أنّ:

$$CL = \bar{p} = \frac{\sum_{i=1}^{m} d_i}{\sum_{i=1}^{m} n_i} \qquad \text{:خط الوسط} \qquad \textbf{(9.32)}$$

حيث m عدد العينات المأخوذة. أمّا حدي الضبط العلوي والسفلي فيأخذان الصيغة التالية:

$$UCL_i = \overline{p} + 3\sigma_{\hat{p}} = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n_i}} \qquad \textbf{(9.33)}$$

$$LCL_i = \overline{p} - 3\sigma_{\hat{p}} = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n_i}} \qquad \textbf{(9.34)}$$

وهنالك تعديل على الطريقة السابقة وهو استخدام متوسط الأحجام المختلفة للعينات $\overline{n} = \frac{\sum_{i=1}^{m} n_i}{m}$ حيث m عدد العينات، بدلاً من n_i في كل من العلاقتين السابقتين (9.33) و (9.34). لنحصل على علاقات جديدة لحدود الضبط العليا والدنيا للنسب.

9-10 خرائط أعداد العيوب (c − charts):

يَرمز الحرف c لأعداد العيوب أو عدم التطابق مع المواصفات المطلوبة (non conformities) في منتجات العينة الواحدة المأخوذة من العملية الإنتاجية وعادةً ما تؤخذ أحجام العينات في مثل هذه الحالة صغيرة $n \geq 1$ مثلاً فإذا كان عدد العيوب في العينة الصغيرة كبيراً فإنّ ذلك مؤشر ينذر بالخطر وعليه يمكن تكوين نوعين من خرائط C يعتمد أحدهما على عدد العيوب في الوحدة الإنتاجية الواحدة. بينما يعتمد النوع الثاني على متوسط العيوب $\overline{c}$ في العينة الواحدة (أي عدد العيوب للوحدة الواحدة)، علاوة على افتراض ثبوت حجم العينة بحيث تتناسب مع توزيع بواسون حيث يتساوى المتوسط مع التباين وهذا يعني أنّ عدد فرص حصول عدم التطابق مع المواصفات (العيوب) كبير جداً وقد يؤول إلى ∞ بينما احتمال وقوع أو حصول أي عيب في أي موقع أو مرحلة من المراحل صغير جداً وثابت وعلى هذا الأساس نفترض أنّ:

$$x = 0, 1, 2, \ldots$$

هو المتغير العشوائي الذي يمثل عدد العيوب في الوحدة الإنتاجية الواحدة أو مجموعة من الوحدات الإنتاجية وأنّ العدد c يمثل كلاً من المتوسط μ والتباين σ^2 المعلومين في هذا التوزيع (توزيع بواسون $\mu = c = \sigma^2$) حيث:

$$P(x) = \frac{e^{-c} c^x}{x!}$$

ومنه يمكن تكوين حدود الضبط كالتالي:

$$CL = c$$

$$UCL = c + 3\sqrt{c} \quad (9.35)$$

$$LCL = c - 3\sqrt{c} \quad (9.36)$$

وعندما يكون $LCL < 0$ سالباً فإننا نستعيض عنه بالصفر $(LcL = 0)$. وفي حالة عدم معرفة كل من المتوسط μ والتباين σ^2 للمجتمع الإحصائي (الإنتاج قيد الدراسة) أي عند عدم معرفة c نستطيع استخدام التقدير النقطي للعدد c وذلك باستخدام $\overline{c}$ (متوسط العيوب في العينات) أي أنّ حدود الضبط في مثل هذه الحالات ستكون كالتالي:

$$CL = \overline{c}$$

$$UCL = \overline{c} + 3\sqrt{\overline{c}} \quad (9.37)$$

$$LCL = \overline{c} - 3\sqrt{\overline{c}} \quad (9.38)$$

وتعتبر خرائط عدد العيوب c أكثر وضوحاً وموثوقية من خرائط النسبة p نظراً لتعددية أنواع وأسباب العيوب وتشخيصها.

مثال (5): لوحظت عيوب الإنتاج لأحدى الشركات المنتجة للدارات الكهربائية الالكترونية من خلال 26 عينة متتالية زمنياً تحتوي كل منها على مائة وحدة إنتاجية وسجلت أعداد العيــوب في الجدول التالي:

الجدول (9-6)

عدد العيوب	تسلسل العينات	عدد العيوب	تسلسل العينات
19	14	21	1
10	15	24	2
17	16	16	3
13	17	12	4
22	18	15	5
18	19	5	6
39	20	28	7
30	21	20	8
24	22	31	9
16	23	25	10
19	24	20	11
17	25	24	12
15	26	16	13

حيث يكون مجموع العيوب = 516 عيباً

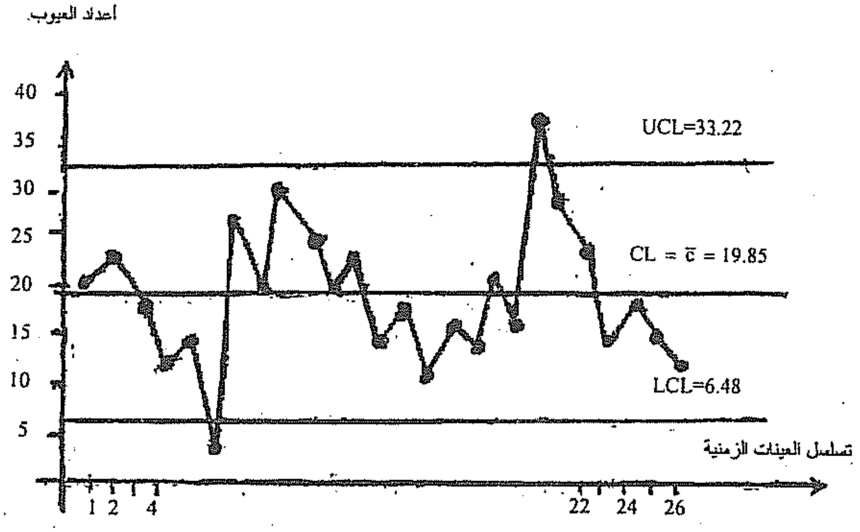
ومنها نجد أنّ حدود الضبط التجريبية (Trial cont. limits) لخارطة c تحسب باستخدام التقدير المتوسط $\bar{c}$ نظراً لعدم معلوميه كل من μ،σ^2 أي أنّ:

$$CL = \bar{c} = \frac{516}{26} = 19.85$$

$$UCL = \bar{c} + 3\sqrt{\bar{c}} = 19.85 + 3\sqrt{19.85} = 33.22$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}} = 19.85 - 3\sqrt{19.85} = 6.48$$

وبهذا تكوّن خارطة c لأعداد العيوب كما في الشكل التالي:



الشكل (9-17): خارطة c لأعداد العيوب في المثال (5)

ويلاحظ من الخارطة السابقة (الشكل (9-17)) بأنّ هنالك نقطتين خارج حدود الضبط التجريبية تمثلان أعداد العيوب في كل من العينتين 6 ، 20 وبعد دراسة كل من العينتين وجد أنّ الخلل إداري ناتج عن تبديل مراقب خط الإنتاج وعدم معرفته برصد العيوب في العينة 6 بينما كان الخلل في الحالة الثانية حراري مما أدى إلى خلل في اللحام الموصل للكهربائية وبمعرفة هذه الأسباب وتصليح الخلل نستطيع إهمال أو غض النظر عن كل من العينتين 6 ، 20 وتكوين خارطة c الجديدة وبحدود الضبط المعدّلة كالتالي:

(حيث أصبح عدد العينات 24 وأصبح مجموع العيوب هو 472 بعد حذف العينتين 6 ، 20)

خط الوسط المعدّل $CL = \bar{c} = \frac{472}{24} = 19.67$

حدي الضبط العلوي والسفلي المعدّلين هما:

$$UCL = \bar{c} + 3\sqrt{\bar{c}} = 19.67 + 3\sqrt{19.67} = 32.97$$

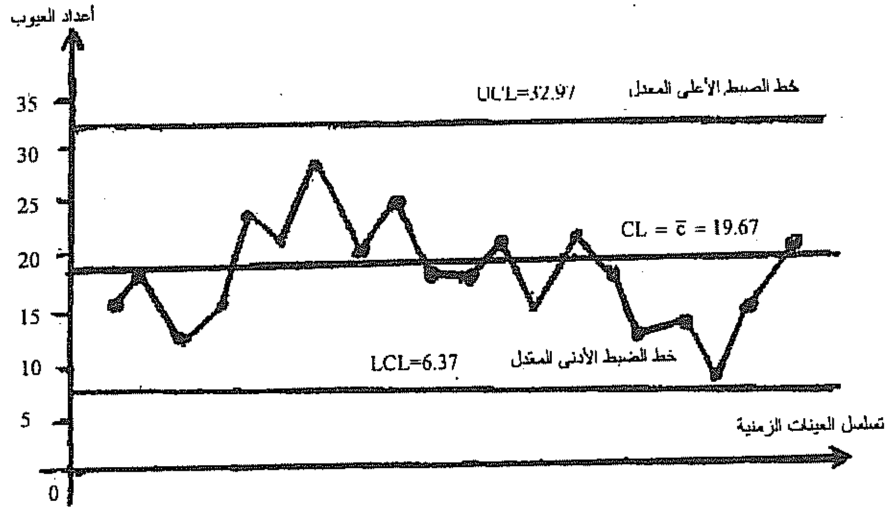
$$LCL = \bar{c} - 3\sqrt{\bar{c}} = 19.67 - 3\sqrt{19.67} = 6.37$$

وتعتبر هذه الحدود المعدلة مقياساً معتمداً لضبط المنتجات الجديدة بعد مرحلة التصليح والصيانة.

وبأخذ مجموعة جديدة من العينات بعد التصحيح متكونة من 20 عينة وبصورة متتالية، تمّ رصد العيوب فيها كما في الجدول التالي:

الجدول (9-7)

عدد العيوب	تسلسل العينة	عدد العيوب	تسلسل العينة	عدد العيوب	تسلسل العينة
19	15	20	8	16	1
12	16	25	9	18	2
14	17	19	10	12	3
9	18	18	11	15	4
20	19	21	12	24	5
21	20	16	13	21	6
		22	14	28	7



الشكل (9-18): خارطة العيوب المعدّلة والمقبولة للمثال (5)

وبمطالعة خارطة العيوب شكل (9-18) للمجموعة الجديدة نلاحظ عدم وجــود أي نقطــة أو قيمة $\overline{c}$ خارج حدود الضبط على الرغم من تشتتها الملحوظ حول الوسط وبهذا نســتنتج بــأنّ العملية الإنتاجية منضبطة أو محكمة.

9-11 خرائط أعداد العيوب u للوحدات الإنتاجية (u-charts):

يعتمد تكوين خرائط عدد عيوب الوحدات الإنتاجية $(u-\text{charts})$ على متوسط عدد العيــوب للوحدة الإنتاجية الواحدة في العينة فإذا كان مجموع عدد العيوب هو c في عينــة ذات حجــم ثابت n فإنّ متوسط العيوب للوحدة الواحدة: $u=\frac{c}{n}$ حيث U عبارة عن متغير عشوائي لتوزيع بواسون أيضاً وحيث يعتبر مجموعاً خطياً لـ n من متغيرات بواسون العشوائية وعليــه نجد أنّ حدود الضبط والمراقبة في هذه الخرائط تأخذ القيم التالية:

$$CL=\overline{u}=\frac{\sum_{i=1}^{m}u_i}{m} \qquad \text{خط الوسط:} \quad (9.39)$$

حيث m عدد العينات العشوائية، $\overline{u}$ متوسط متوسطات العيوب للوحدة الواحدة.

$$UCL=\overline{u}+3\sqrt{\overline{u}/n} \qquad \text{الحد الأعلى للضبط:} \quad (9.40)$$

$$LCL=\overline{u}-3\sqrt{\overline{u}/n} \qquad \text{الحد الأدنى للضبط:} \quad (9.41)$$

مثال (6): يرغب أحد مهندسي الضبط والمراقبة لشركة خاصة لتصنيع أجهزة الكمبيوتر فــي تكوين خارطة ضبط لمراقبة عدد العيوب للوحدة الإنتاجية الواحدة. أخذت 20 عينة عشــوائية متتالية ذات حجم ثابت $n=5$ وسجلت أعداد العيوب من هذه العينات كما في الجدول التالي:

الجدول (9-8)

التسلسل i	الحجم n	c_i مجموع العيوب	$u_i = \frac{c_i}{n}$ متوسط العيوب للوحدة	التسلسل i	الحجم n	c_i مجموع العيوب	$u_i = \frac{c_i}{n}$ متوسط العيوب للوحدة
1	5	10	2.0	11	5	9	1.8
2	5	12	2.4	12	5	5	1.0
3	5	8	1.6	13	5	7	1.4
4	5	14	2.8	14	5	11	2.2
5	5	10	2.0	15	5	12	2.4
6	5	16	3.2	16	5	6	1.2
7	5	11	2.2	17	5	8	1.6
8	5	7	1.4	18	5	10	2.0
9	5	10	2.0	19	5	7	1.4
10	5	15	3.0	20	5	5	1.0

حيث نجد أنّ مجموع العيوب = 193 ومجموع متوسطات العيوب للوحدات = 38.6

ومن المعلومات أعلاه نجد أنّ:

خط الوسط:

$$CL = \bar{u} = \frac{\sum_{i=1}^{20} u_i}{20} = \frac{38.60}{20} = 1.93$$

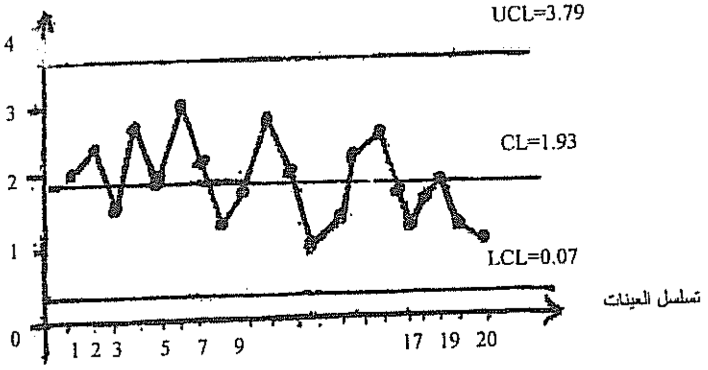
وأنّ: الحد الأعلى:

$$UCL = \bar{u} + 3\sqrt{\bar{u}/n} = 1.93 + 3\sqrt{1.93/5} = 3.79$$

الحد الأدنى:

$$LCL = \bar{u} - 3\sqrt{\bar{u}/n} = 1.93 - 3\sqrt{1.93/5} = 0.07$$

حيث نحصل على الخارطة التالية:



الشكل (9-19): خارطة العيوب للوحدة الواحدة للمثال (6)

في كثير من حالات تعداد العيوب نصطدم بالواقع ونلاحظ عدم إمكانية توحيد أحجام العينات كما في المثال السابق. فمثلاً في مراقبة لشركات النسيج أو تصنيع الورق نجد عدم إمكانية توحيد أحجام العينات المأخوذة لأنّ لفات النسيج أو لفات الورق تختلف في أطوالها وفي عرضها فعند اختيارنا عشر لفات من النسيج نجد أنّ مساحتها الكلية تختلف عن مساحة عشر لفات أخرى مختارة عشوائياً من المصنع وكذلك بالنسبة للورق وفي مثل هذه الحالات نستخدم أسلوب الأحجام المتغيرة وغير الثابتة وعلى ضوء استخدام هذا الأسلوب نجد أنّ حدود الضبط العليا والدنيا تتغير هي الأخرى وليست ثابتة أحياناً أو متغيرة لبعض العينات وثابتة لمجموعة أخرى منها.

مثال (7): في مراقبة لإحدى شركات النسيج تم أخذ عشرة لفات متتالية من النسيج وتم رصد عدد العيوب على كل 50 متر مربع فيها وعلى ضوء المساحة الكلية لكل لفة أو ربطة من الوحدات المنتجة، كما في الجدول التالي لتكوين وبناء خارطة العيوب في الوحدة الواحدة أو اللفة الواحدة ($u-charts$).

الجدول (9-9)

تسلسل لفات النسيج i	المساحة الكلية بالمتر المربع	مجموع العيوب	تعداد المساحات في اللفة الواحدة	متوسط العيوب للوحدة الواحدة u_j
1	500	14	10	1.4
2	400	12	8	1.5
3	650	20	13	1.54
4	500	11	10	1.10
5	475	7	9.5	0.74
6	500	10	10	1.00
7	600	21	12	1.75
8	525	16	10.5	1.52
9	600	19	12	1.58
10	625	23	12.5	1.84
		مجموع العيوب الكلي: 153	مجموع وحدات المساحات المفحوصة: 107.50 وحدة	

ومن الجدول السابق نجد أنّ:

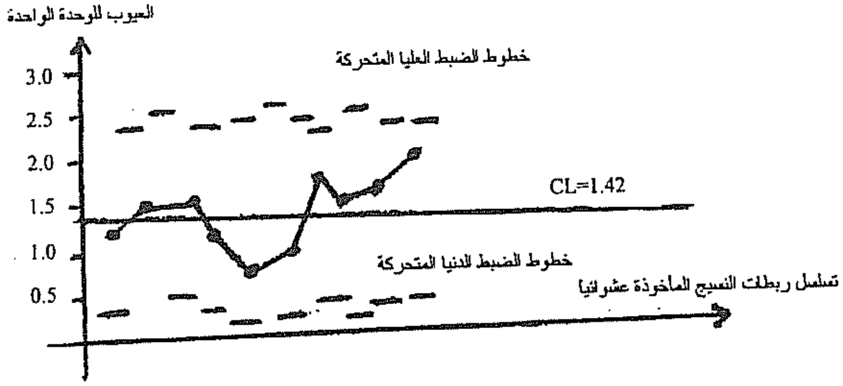
خط الوسط الثابت $CL = \bar{u} = \frac{153}{107.5} = 1.42$

وعلى ضوء تغير أعداد المساحات المفحوصة أو المراقبة في كل لفة من النسيج يتوجب استخدام تلك الأعداد n_i بدلاً من n (عدد المفردات في العينة) لكل لفة من النسيج بصورة منفردة لتكوين خطوط الضبط والمراقبة العليا والدنيا للبيانات السابقة وكما يلي:

الجدول (9-10)

تسلسل لفات النسيج	أعداد المساحات (50 م²) في كل لفة n_i	$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n_i}}$	$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n_i}}$
1	10	2.55	0.29
2	8	2.68	0.16
3	13	2.41	0.43
4	10	2.55	0.29
5	9.5	2.58	0.26
6	10	2.55	0.29
7	12	2.45	0.39
8	10.5	2.52	0.32
9	12	2.45	0.39
10	12.5	2.43	0.41

حيث نحصل على الشكل التالي:



الشكل (9-20): خارطة العيوب للوحدة الواحدة لأحجام متغيرة

9-12 توزيعات باريتو (pareto distribution):

وهي عبارة عن أشكال إحصائية وصفية تكرارية لعيوب الإنتاج وتعطى على أشكال مختلفة كمضلعات تكرارية أو نقطية حيث يتم فيها تحديد نوعية العيوب وعدم المطابقة مع المواصفات وتكراراتها أو نوعية العيوب ونسب تلك التكرارات وتعطى على شكل التكرارات المجتمعة الصاعدة عادة. وبديهي أنّ تشخيص نوعية العيوب يسّهل من عمليات الصيانة وتطوير الإنتاج سواءً كانت هذه العيوب عددية أو وصفية وعلى أساس أن الكثير من الأعطال هي صفة للنادر أو القليل من أنواع هذه الأعطال، يتم توزيع باريتو كما في الشكل التالي حيث نبدأ بالتكرارات الكبيرة عادة للأعطال وهكذا بصورة تنازلية:

الجدول (11-9)

التوزيع التكراري النسبي المتجمع الصاعد لأنواع الأعطال في عملية تصنيع قطـع الـدارات الكهربائية لباريتو

التكرار النسبي المتجمع	التكرار النسبي	التكرار المتجمع	التكرار		رقم أو نوع الخلل	التسلسل
40.82	40.82	40	40	**	لحام غير كافي	1
61.22	20.41	60	20	********************	لحام بارد	2
68.37	7.14	67	7	*******	لحـــام مفتوح	3
74.49	6.12	73	6	******	قطع غير مضبوطة	4
79.59	5.10	78	5	*****	لحام	5
82.65	3.06	81	3	***	مؤشـــر الاختبار	6
85.71	3.06	84	3	***	....	7
88.78	3.06	87	3	***	....	8
90.82	2.04	89	2	**	زيـــادة قطع	9
92.86	2.04	91	2	**	قطـــع معطوبة	10
94.90	2.04	93	2	**	قطـــع مفقودة	11
95.92	1.02	94	1	*	أســلاك غيـــر مضبوطة	12
96.94	1.02	95	1	*	خـــتم العامل	13
97.96	1.02	96	1	*	أختـــام مفقودة	14
98.98	1.02	97	1	*	لحـــام قصير	15
100.0	1.02	98	1	*	... تالفة	16

ويلاحظ من جدول باريتو السابق بأنّ الخلل الكبير هو أن اللحام في هذه الدارات غير كـــافٍ وعليه يتوجب التركيز على كفايتها إمّا بالصيانة أو باستبدال العامل ...إلخ أمّا الخلـــل الكبيـــر الثاني فهو في اللحام البارد وعليه يتوجب الصيانة واستبدال العامل أو المواد الأولية ...إلـــخ وبنفس الأسلوب يتم صيانة الأعطال واحداً بعد الآخر وعلى الترتيب النسبي لضمان الجودة.

وتعتبر طريقة باريتو لرصد ومراقبة العيوب من الطرق التقنية الهامة التي تعتمد ليس علـــى تعداد العيوب أو نسبها بل على نوعيتها أيضاً لتحديد أولوية الصيانة وضمان الجودة.

الجدول (12-9)

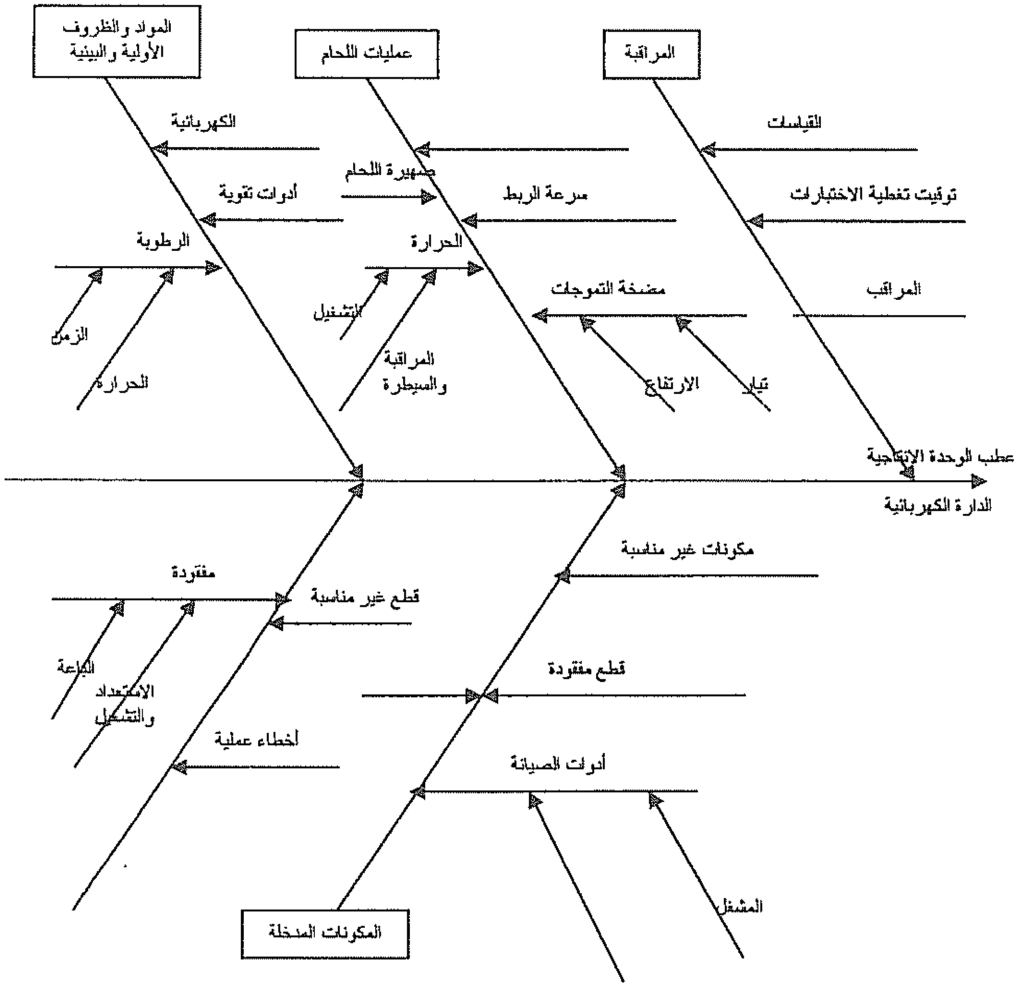
جدول العيوب المصنفة بأرقام القطع وتكراراتها ونسبهاالمئوية ورموز الأعطال لباريتو

رقم القطعة التكرار النسبي نسب السطر نسب العمود	المكونات المفقودة	المكونات التالفة	المكونات الزائدة	رمز التالف	رمز الصنف	رمز الصنف التالف	لحام ناقص	لحام مفتوح DEWE	لحام بارد
0001285	1 -1.02 1.41 50.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 1.02 1.41 100.00	0 0.00 0.00 0.00	5 5.10 7.04 71.43	20 20.41 28.17 100.00
0001481	1 1.02 3.70 50.00	2 2.04 7.41 100.00	2 6.12 22.22 100.00	6 3.06 11.11 100.00	3 0.00 0.00 100.00	0 0.00 0.00 0.00	1 1.02 3.70 100.00	2 2.04 7.41 28.57	0 0.00 0.00 0.00
0006429	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00
المجموع	2 2.04	2 2.04	2 2.04	6 6.12	3 3.06	1 1.02	1 1.02	7 7.14	20 20.41

رقم القطعة، التكرار، التكرار النسبي، نسب السطر، نسب العمود	لحام غير كافي	لحام	ختم مفقود	رمز التالف ختم العامل ID	إشارة الفحص أبيض M	إشارة الفحص إشارة EC	أسلاك غير مضبوطة S	وحدات جيدة (s)	المجموع
0001285	40 40.82 56.34 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	2 2.04 2.82 66.67	1 1.02 1.41 33.33	1 1.02 1.41 100.00	0 0.00 0.00 0.00	72.5
0001481	0 0.00 0.00 0.00	5 5.10 18.52 100.00	1 1.02 3.70 100.00	1 1.02 3.70 100.00	1 1.02 3.70 33.33	2 2.04 7.41 66.67	0 0.00 0.00 0.00	0 0.00 0.00 0.00	27.5
0006429	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0.0
المجموع	40 40.82	5 5.10	1 1.02	1 1.02	3 3.06	3 3.06	1 1.02	0 0.00	0 100.00

9-13 مخططات الأسباب وتأثيراتها (Cause and effect diagram):

تعتبر مخططات الأسباب وتأثيراتها من التقنيات الهامة في تحليل عدم التطابق مع المواصفات المطلوبة للوحدات الإنتاجية وهي مخططات وصفية غير عددية تسمى بالهياكل السمكية أو أشكال إيشيكاوا نسبة إلى كورو إيشيكاوا. وتستخدم مخططات الأسباب وتأثيراتها لتوضيح أنواع الأعطال المختلفة لوحدات الإنتاج ومصادرها وعلاقاتها الداخلية فيما بينها، وتعتبر هذه المخططات كلوحة بيانات مفيدة لكل من العاملين والمهندسين الصناعيين ومدراء الإنتاج ومراقبي ومهندسي الجودة ومشاكلها. وبديهي أنّ جودة الهياكل السمكية هذه تعكس مستوى الفهم التقني للعملية الإنتاجية وأسرارها وهنالك العديد من الأساليب لرسم هذه المخططات أو الهياكل حيث ترسم هذه المخططات وفقاً لأولويات الجودة في العملية الإنتاجية كالتركيز مثلاً على المواد الأولية والعمال ووسائل الإنتاج وقطع الصيانة ...إلخ. وعادةً ما ترسم هذه المخططات وفقاً لتدفق المواد الأولية بدأ من اليسار إلى اليمين خلال العملية الإنتاجية وكما في المخطط التالي حول عملية إنتاج الدارات الكهربائية في المثال (5) السابق:



شكل (9-21): مخطط إيشيكاوا للأسباب وتأثيراتها للأعطال في الوحدات الإنتاجية

والحقيقة إنّ هذه المخططات تعتبر خلاصة إدارية هامة تستوحى عادةً من خلال جلسات كافة العاملين ضمن خط الإنتاج لدراسة وتحليل الأسباب والآثار الناتجة عن الأعطال في العملية الإنتاجية ومن ثم يتم اتخاذ القرار المناسب للصيانة والتطوير أو التحسين لكل مراحل العملية الإنتاجية بما فيها جودة الإنتاج.

9-14 قابلية أو قدرة العملية الإنتاجية (process capability):

تعتبر قابلية أو قدرة العملية الإنتاجية وحساباتها إحدى الوسائل الهامة أيضاً لتحسين الجودة وتطوير الإنتاج ونعني بالقابلية أو القدرة هو إمكانية العملية الإنتاجية على مطابقة المواصفات والشروط المفروضة والمحددة مسبقاً على العملية الإنتاجية (specification limits) وهي تختلف عن حدود الضبط المحسوبة سابقاً.

ولحساب ما إذا كانت العملية الإنتاجية على درجة من القدرة والقابلية لمطابقة المواصفات والشروط المفروضة مسبقاً يتم حساب ما يسمى بمعامل أو دليل القدرة (process capability index) والذي يُرمز له عادةً بالرمز PCI ويُعرَّف على أنه النسبة بين ما يسمى بمدى التحمل (Tolerance)، وهو الفرق بين الحدين الأعلى والأدنى للشروط المفروضة مسبقاً، إلى مدى الانحرافات المعيارية الستة للعملية الإنتاجية المستخدمة سابقاً في خرائط المراقبة أي الفرق بين الحدين الأعلى والأدنى المحسوبين في خرائط المراقبة وبعبارة رياضية يعرف معامل أو دليل القدرة PCI بالعلاقة التالية:

$$PCI = \frac{USL - LSL}{UCL - LCL} = \frac{USL - LSL}{6\sigma} \qquad (9.42)$$

ومن الواضح استقراء العلاقات التالية:

$PCI > 1$ يعني أنّ مدى تشتت المواصفات المشروطة أكبر وأوسع من مدى تشتت متوسطات العينات أو مقاييس الجودة عموماً في العملية الإنتاجية المنضبطة. وهذا يعني وقوع المنتجات ضمن المواصفات المشروطة وبمعنى آخر نستنتج بأنّ العملية الإنتاجية ذات قدرة أو قابلية تتطابق مع أو تفوق الشروط المفروضة للجودة.

$PCI = 1$ يعني تطابق أو تساوي مدى تشتت المواصفات مع مدى تشتت مقاييس الجودة في العملية الإنتاجية. وهذا يعني أيضاً أنّ العملية الإنتاجية ذات قدرة تتطابق تماماً ولا تفوق المواصفات والشروط المفروضة للجودة.

$PCI < 1$ يعني أنّ مدى تشتت المواصفات المشروطة أصغر أو أضيق من مدى تشتت مقاييس الجودة في العملية الإنتاجية المنضبطة. أي أن الحدين الأعلى والأدنى للضبط المحسوبين ضمن مخطط الضبط والمراقبة يقعان خارج المواصفات أو خارج الحدين الأعلى والأدنى المشروطين وبهذا سنجد بأنّ هنالك عدداً من هذه المنتجات خارج المواصفات المشروطة على الرغم من أنها داخل حدود الضبط المحسوبة للعملية الإنتاجية وحسب خارطة

الضبط التي تصف العملية الإنتاجية. كما تعني أيضاً بأنّ قدرة العملية الإنتاجية ليست ضمن الشروط المفروضة والمطلوبة أو بعبارة أخرى يُقال عن العملية الإنتاجية بأنها غير قادرة على مطابقة المواصفات المطلوبة ويتوجب عندئذٍ إجراء التعديل أو الصيانة والتطوير للمطابقة مع الشروط.

ففي المثال (1) لو افترضنا أن الشروط المفروضة على حدي الضبط الأعلى والأدنى لمتوسطات الإنتاج هو: $\mu \pm 3\sigma = 74 \pm 0.03$ أي أنّ: $LSL = 73.970$ وأنّ: $USL = 74.030$ بينما تشير حسابات المثال رقم (1) إلى أنّ حدي الضبط الأعلى والأدنى هما $LCL = 73.988 > LSL$ وأنّ: $UCL = 74.014 < USL$ أي أنّ هنالك اختلافاً بين حدي الضبط المفروضين وحدي الضبط المحسوبين أو الواقعيين (UNTL, LNTL) وأنّ هنالك نوعاً من الإزاحة بينهما مما يتوجب عليه حساب دليل أو معامل القابلية PCI لمعرفة كم من المتوسطات خار شروط المطلوبة على الرغم من أنها داخل حدود الضبط المحسوبة أو الواقعية للعملية الإنتاجية وعليه نجد أنّ:

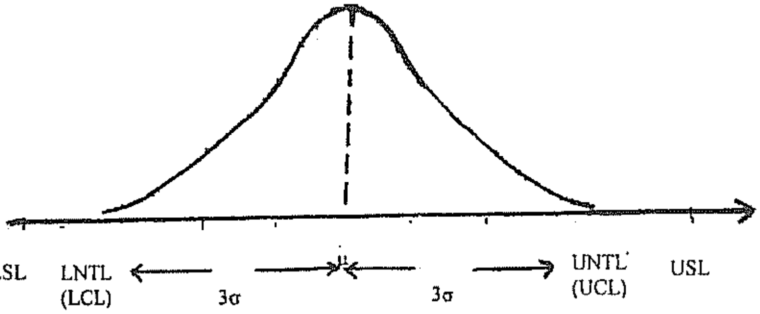
$$PCI = \frac{USL - LSL}{3\sigma - (-3\sigma)} = \frac{74.03 - 73.97}{6(0.0099)} = \frac{0.06}{0.0594} = 1.01$$

وهذا العدد PCI=1.01 يكاد يساوي الواحد مما يعني أنّ قدرة العملية الإنتاجية تكاد تتطابق مع المواصفات والشروط المفروضة أو أنّ مدى التحمل الواقعي للعملية الإنتاجية يكاد يتساوى مع الحدود والشروط المطلوبة أي أنّ هنالك عدداً ضئيلاً جداً ولا يعتد به سيكون خارج المواصفات والحدود أو الشروط المفروضة.

وللحصول على نسبة أفضل لأعداد أو نسب الوحدات المعطوبة والتالفة نستخدم الدرجة المعيارية على افتراض أنّ متوسطات حلقات المكابس للسيارات تتوزع توزيعاً طبيعياً $N(\mu, \sigma)$ وعليه نجد أنّ احتمالية العيوب أي احتمالية وقوع الوحدات المنتجة في منطقة الرفض على جهتي التوزيع هو احتمال وقوع متوسط المنتج دون الحد الأدنى LSL وفوق الحد الأعلى USL كما يلي:

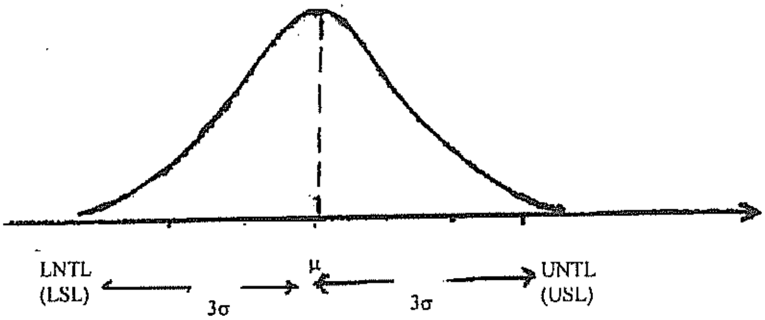
$$\hat{p} = p(X < 73.970) + p(X > 74.03)$$
$$= p\left(\frac{73.970 - 74.001}{0.0099}\right) + \left(1 - p(\frac{74.03 - 74.001}{0.0099})\right)$$
$$= p(-3.13) + 1 - p(2.93) = 0.00087 + 1 - 0.99831 = 0.00256$$

ويعني هذا أنّ هنالك 256 وحدة منتجة معطوبة من بين كل مائة ألف وحدة منتجة أي أنّ نسبة حلقات المكابس غير المطابقة للمواصفات والشروط المطلوبة هـي 0.256٪ خـارج حـدي الضبط LSL ، USL. ويمكن توضيح ذلك بيانياً بالأشكال التالية:



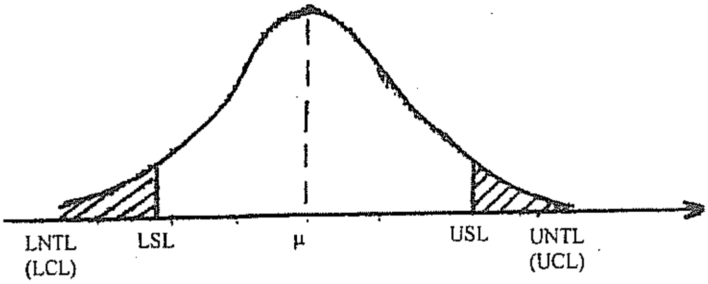
PCI > 1 -(a)

حدود الضبط المحسوبة ضمن المواصفات المشروطة والقدرة تفوق الشروط



PCI = 1 -(b)

تطابق بين الحدود الطبيعية المحسوبة للعملية الإنتاجية مع الحدود المشروطة والقدرة مطابقة للشروط والمواصفات



PCI < 1 -(c)

العملية الإنتاجية ذات قدرة غير مطابقة للشروط المفروضة وهنالك عدد من الوحدات الإنتاجية خارج الحدود المشروطة على الرغم من أنها داخل حدود الضبط الطبيعية

شكل (22-9)

ومن الجدير بالذكر بأنّ الدليل PCI يمكن أن يعرّف ويحسب لكل منطقة من مناطق خارطة الضبط الستة A ، B ، C حيث نلاحظ من خرائط الضبط بأنّ:

$$PCI_c = \frac{USL - LSL}{2\sigma}$$

$$PCI_B = \frac{USL - LSL}{4\sigma}$$

$$PCI_A = PCI = \frac{USL - LSL}{6\sigma}$$

مثال (8): استلمت شركة لتصنيع أقراص كمبيوتر (3.5 inch) شكاوى عن كثير من الأقراص لا تخزّن البيانات بصورة مضبوطة، وعليه قررت الإدارة تكوين خارطة لضبط العملية فأخذت عينة عشوائية كل ساعة، كل منها ذات حجم n = 200 قرص حيث يتم فحص كل قرص من الأقراص لمعرفة ما إذا كان معطوباً أم لا وسجلت النتائج بالجدول التالي:

جدول (9-13)

العينات	عدد الوحدات المعطوبة	العينات	عدد الوحدات المعطوبة	العينات	عدد الوحدات المعطوبة
1	19	14	10	27	9
2	5	15	18	28	10
3	16	16	20	29	10
4	20	17	13	30	18
5	6	18	6	31	15
6	12	19	8	32	16
7	18	20	3	33	5
8	6	21	8	34	14
9	13	22	7	35	3
10	15	23	4	36	10
11	10	24	19	37	19
12	6	25	3	38	13
13	7	26	19	39	19
				40	9

ولتكوين جدول نسبة المعطوب (p-chart) نحسب النسب أولاً لكل عينة من العينات:

حيث $p_i = \frac{d_i}{200}$

ومنها نحسب متوسط النسب: $\overline{p} = \frac{\sum_{i=1}^{40} p_i}{40} = 0.5762$

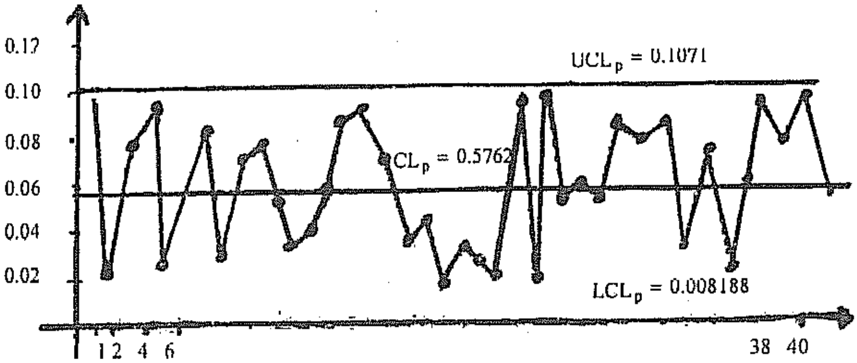
وعليه فإنّ حدي الضبط الأعلى والأدنى

$$UCL_p = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0.5762 + 3\sqrt{\frac{0.5762(1-0.5762)}{200}} = 0.1071$$

$$LCL_p = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0.5762 - 3\sqrt{\frac{0.5762(1-0.5762)}{200}} = 0.008188$$

وبنفس الأسلوب يمكن حساب حدود ضبط المناطق الأخرى الستة حيث نأخذ $k = \pm 1, \pm 2$

ونحصل على الشكل التالي:



شكل (9-23)

حيث يُلاحظ بأنّ العملية الإنتاجية منضبطة أو محكمة على الرغم من تشتت النسب حول CL

ويلاحظ أيضاً من الشكل بأنّ العملية الإنتاجية منضبطة أو محكمة.

9-7 خرائط ضبط المدى والمتوسط المشتركة ($\bar{X}$ and R charts):

يمكننا الاستفادة من حسابات خرائط الجودة والمراقبة لكل من المدى R والمتوسط $\bar{X}$ لتكوين خرائط مشتركة بينها. فلو استخدمنا $\bar{\bar{x}}$ كتقدير لمتوسط المجتمع μ واستخدمنا $\hat{\sigma}=\frac{\bar{R}}{d_2}$ كتقدير غير منحاز لانحراف المجتمع المعياري σ فإنّ حدود خرائط المدى والمتوسط المشتركة تأخذ القيم التالية (حسب الاختيار المتعارف 3σ):

$$UCL=\bar{\bar{x}}+3\frac{\hat{\sigma}}{\sqrt{n}}=\bar{\bar{x}}+\frac{3}{\sqrt{n}}\frac{\bar{R}}{d_2}=\bar{\bar{x}}+A_2\bar{R} \qquad \textbf{(9.10)}$$

الحد الأوسط (خط الاعتدال) $CL=\bar{\bar{x}}$

$$LCL=\bar{\bar{x}}-3\frac{\hat{\sigma}}{\sqrt{n}}=\bar{\bar{x}}-\frac{3}{\sqrt{n}}\frac{\bar{R}}{d_2}=\bar{\bar{x}}-A_2\bar{R} \qquad \textbf{(9.11)}$$

حيث: $A_2=\frac{3}{d_2\sqrt{n}}$ وهي قيم محسوبة في جداول خاصة أيضاً.

وأنّ: $\hat{\sigma}=\frac{\bar{R}}{d_2}=\frac{0.023}{2.326}=0.0099$

مثال (3): استخدم البيانات في المثال (1) لتكوين خارطة المدى والمتوسط المشتركة.

الحل: عرفنا من المثالين (1) و (2) بأنّ: $\bar{\bar{x}}=74.001=CL$, $\bar{R}=0.023$

ولحساب حدود الضبط العليا والدنيا نجد من الجداول بأنّ:

$$A_2=\frac{3}{d_2\sqrt{n}}=\frac{3}{\sqrt{5}\,d_2}=0.577$$

$$UCL_{\bar{x},R}=\bar{\bar{x}}+A_2\bar{R}=74.001+(0.577)(0.023)=74.014$$

$$LCL_{\bar{x},R}=\bar{\bar{x}}-A_2\bar{R}=74.001-(0.577)(0.023)=73.988$$

9-8 خرائط ضبط الانحراف المعياري والمتوسط المشتركة ($\bar{X}$ and S charts):

عندما تكون أحجام العينات كبيرة ولنقل مثلاً $n > 10$ نجد أنّ عملية تقدير σ باستخدام المدى R تفقد كفاءتها نظراً لضياع الكثير من البيانات بين حدي المدى الواسع R وعلى هذا الأساس يفضل في مثل هذه الحالة استبدال خرائط $\bar{X}$ و R بخرائط $\bar{X}$ و S.

وعلى افتراض أنّ تباين العملية الإنتاجية σ^2 غير معلوم لذا نستخدم التقدير المستقل وغير المنحاز للعينة.

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad (9.12)$$

ولكن S ليس تقديراً غير منحاز للانحراف المعياري σ وعلى هذا الأساس نجد أنّ S عبارة عن تقدير لوسيط مثل $c_4\sigma$ أي أنّ $E(s) = c_4\sigma$ حيث c_4 معامل التقدير:

$$c_4 = \left(\frac{2}{n-1}\right)^{\frac{1}{2}} \frac{\Gamma(n/2)}{\Gamma(\frac{n-1}{2})} \qquad (9.13)$$

وهو ثابت محسوب ومجدول يعتمد على حجم العينة العشوائية n أمّا تباين الانحرافات المعيارية S للعينات فيقدر أو يحسب من العلاقة التالية $\sigma\sqrt{1-c_4^2}$. وهذه المعلومة تستخدم في حساب حدود خرائط ضبط S.

وعلى افتراض أنّ σ معلومة فإنّ حدود الضبط لخرائط S تكتب كالتالي:

$$CL = c_4\sigma \qquad (9.14)$$

$$UCL = c_4\sigma + 3\sigma\sqrt{1-c_4^2} \quad B_6\sigma \qquad (9.15)$$

حيث: $B_6 = c_4 - 3\sqrt{1-c_4^2}$ قيم محسوبة ومجدولة مسبقاً.

$$LCL = c_4\sigma - 3\sigma\sqrt{1-c_4^2} = B_5\sigma \qquad (9.16)$$

حيث: $B_5 = c_4 - 3\sqrt{1-c_4^2}$ قيم مجدولة مسبقاً.

وعند عدم معلومية σ يتم أخذ m من العينات العشوائية كل منها ذات حجم n لتقدير σ وذلك باستخدام متوسط الانحرافات $\bar{S}$ حيث نجد منها أنّ:

$$\bar{S} = \frac{1}{m}\sum_{i=1}^{m} s_i \qquad (9.17)$$

حيث s_i يمثل الانحراف المعياري للعينة رقم i.

ومنها نجد أنّ الإحصائية $\bar{S}/c_4$ تمثل تقديراً غير منحاز للانحراف σ وعليه فإنّ حدود الضبط والمراقبة لخرائط S تحسب بالعلاقات التالية:

$$CL = \bar{S} = c_4\sigma$$
$$UCL = \bar{S} + 3\frac{\bar{S}}{c_4}\sqrt{1-c_4^2} = \bar{S}(1+\frac{3}{c_4}\sqrt{1-c_4^2}) = B_4\bar{S} \qquad (9.18)$$

$$LCL = \bar{S} - 3\frac{\bar{S}}{c_4}\sqrt{1-c_4^2} = \bar{S}(1-\frac{3}{c_4}\sqrt{1-c_4^2} = B_3\bar{S} \qquad (9.19)$$

وعند استخدام $E(\frac{\bar{S}}{c_4}) = \sigma$ فإنّ حدود الضبط في خرائط $\bar{x}$ تعرف وتحدد بالتالي:

$$UCL = \bar{\bar{x}} + \frac{3\bar{S}}{c_4\sqrt{n}} = \bar{\bar{x}} + A_3\bar{S} \qquad (9.20)$$

$$CL = \bar{\bar{x}}$$

حيث $A_3=3/c_4\sqrt{n}$ قيم محسوبة ومجدولة.

$$LCL = \bar{\bar{x}} - \frac{3\bar{S}}{c_4\sqrt{n}} = \bar{\bar{x}} - A_3\bar{S} \qquad (9.21)$$

وعلى ضوء ما سبق يمكننا تكوين خرائط ضبط التباين S^2 والتي تعتمد مباشرة على حساب تباين العينات كالتالي:

$$CL = \bar{S}^2$$
$$UCL = \frac{\bar{S}^2}{n-1}\chi^2_{(\alpha/2,n-1)} \qquad (9.22)$$
$$LCL = \frac{\bar{S}^2}{n-1}\chi^2_{(1-\alpha/2,n-1)} \qquad (9.23)$$

9-9 خرائط ضبط النسبة P (P charts):

تعتبر خرائط ضبط النسبة من مجموعة خرائط ضبط المواصفات غير الكمية مثل تالف أو غير تالف، يعمل أولاً يعمل، معيب أو غير معيب، ويُقصد بالنسبة هنا هو عدد المنتجات المعطوبة في العينة إلى عدد المنتجات فيها مثلاً أو عدد المنتجات غير المطابقة للمواصفات إلى عدد المنتجات الكلي في العينة وتستخدم هذه الخرائط عادةً عندما يراد ضبط ومراقبة

أعداد التالف أو المعطوب في العملية الإنتاجية ويتم تكوين هذه الخرائط بنفس الأسلوب السابق ويُراعى فيها عدد العينات $m \geq 25$ وبحجم $n \geq 50$ لضمان تحقق تقارب توزيع النسب من التوزيع الطبيعي.

ومن التعريف السابق نلاحظ بأنّ خرائط النسبة تعتمد بطبيعتها على ما يسمى بالتوزيع الثنائي (Binomial distribution). فلو افترضنا أنّ نسبة التالف أو المعطوب أو غير المطابق للمواصفات الفنية هي $\bar{p}$ وأنّ الوحدات المنتجة مستقلة وغير مترابطة ضمن عملية الإنتاج نجد أنّ كل وحدة من وحدات الإنتاج تمثل قيمة عشوائية لمتغير برنولي ثنائي التوزيع وبالنسبة (أو الوسيط) p. فإذا تم أخذ عينة عشوائية ذات حجم n من الوحدات وبفرض d هو عدد الوحدات غير المطابقة للمواصفات فإنّ D يعتبر متغيراً عشوائياً ثنائي التوزيع $B(n,p)$ ذو وسيطين هما (n,p) وعليه فإنّ:

$$P[D=x]=\binom{n}{x}p^{x}(1-p)^{n-x} \qquad x=0,1,\ldots,n \qquad \textbf{(9.24)}$$

ومن معلوماتنا السابقة حول التوزيع الثنائي وتوزيع برنولي نجد أنّ المتوسط $\mu = np$ وأنّ التباين $S^2 = npq = np(1-p)$. ولو افترضنا أنّ نسبة التالف أو المعطوب في كل عينة هو $\hat{p} = \frac{d}{n}$ فإنّ المتوسط هو $\mu = p$ والتباين هو $\sigma^2_{\hat{p}} = \frac{p(1-p)}{n}$

وعلى ضوء معطيات خرائط ضبط شيوارت السابقة يمكننا صياغة حدود الضبط العليا والدنيا وخط الوسط مفترضين معلومية p مسبقاً كالتالي:

$$UCL = p + 3\sqrt{\frac{p(1-p)}{n}} \qquad \textbf{(9.25)}$$

$$CL = p$$

$$LCL = p - 3\sqrt{\frac{p(1-p)}{n}} \qquad \textbf{(9.26)}$$

حيث اخترنا حدود الضبط ثلاث وحدات انحرافية معيارية كالمتعارف. وعند عدم معرفة نسبة المعطوب أو التالف (p) عموماً فيمكن تقديرها وحسابها من العينات المأخوذة عشوائياً بعدد $m \geq 25$ على الأقل وكل منها بحجم $n \geq 50$ وبديهي أن يؤدي زيادة حجم العينة إلى دقة في نسبه المعطوب أو التالف من المنتجات في العينة الواحدة. ولو افترضنا في هذه الحالة أن النسبة لكل عينة هي:

$$\hat{p}_i = \frac{d_i}{n}$$

حيث d_i عدد المعطوب أو التالف في العينة رقم i $(i=1,2,\ldots,m)$
وعليه يمكن حساب متوسط النسب في هذه العينات:

$$\overline{p}=\frac{\sum_{i=1}^{m}d_i}{mn}=\frac{\sum_{i=1}^{m}\hat{p}_i}{m} \qquad (9.27)$$

حيث المتوسط أو الإحصائية $\overline{p}$ تعتبر تقديراً غير منحازاً للنسبة p غير المعلومة للمجتمع الإحصائي أي: $\mu=E(\overline{p})=p$ وكذلك $\sigma_{\hat{p}}=\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$ يعتبر تقديراً للانحراف المعياري للمجتمع أي: $\sigma=E(\sigma_{\hat{p}})$ وعليه يمكننا تكوين خارطة ضبط النسبة p وحساب حدود الضبط العليا والدنيا والوسط (مفترضين عدم معلوميه p) كالتالي:

$$UCL_p=\overline{p}+3\sigma_{\hat{p}}=\overline{p}+3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \qquad (9.28)$$

$$CL_p=\overline{p}$$

$$LCL_p=\overline{p}-3\sigma_{\hat{p}}=\overline{p}-3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \qquad (9.29)$$

مثال (4): في مراقبة ميدانية لمكائن تعبئة الحليب في علب ورقية سميكة لأحدى الشركات الصناعية حيث كان المراد مراقبة وضبط تسرب الحليب من جوانب أو أسفل هذه العلب وتكوين خارطة نسبة المعيب من هذه العلب (أي العلب القابلة للتسريب)، تم أخذ 30 عينة عشوائية من هذه العلب المملوءة كل منها يحتوي على 50 علبة حليب مملوءة على فترات كل نصف ساعة وعلى كافة مراحل عمل المصنع الثلاث. ورصدت النسب كما يلي:

جدول (3-9)

رقم العينة	عدد العلب القابلة للتسريب d_i (المعيبة)	نسبة المعيب في العينة $\hat{p}_i = \frac{d_i}{n}$
1	12	0.24
2	15	0.30
3	8	0.16
4	10	0.20
5	4	0.08
6	7	0.14
7	16	0.32
8	4	0.18
9	14	0.28
10	10	0.20
11	5	0.10
12	6	0.12
13	17	0.34
14	12	0.24
15	22	0.44
16	8	0.16
17	10	0.20
18	5	0.10
19	13	0.26
20	11	0.22
21	20	0.40
22	18	0.36
23	24	0.48
24	15	0.30
25	9	0.18
26	12	0.24
27	7	0.14
28	13	0.26
29	9	0.18
30	6	0.12
	المجموع = 347	$\bar{p} = 0.2313$

الحل: من الجدول المبين أعلاه نلاحظ أنّ:

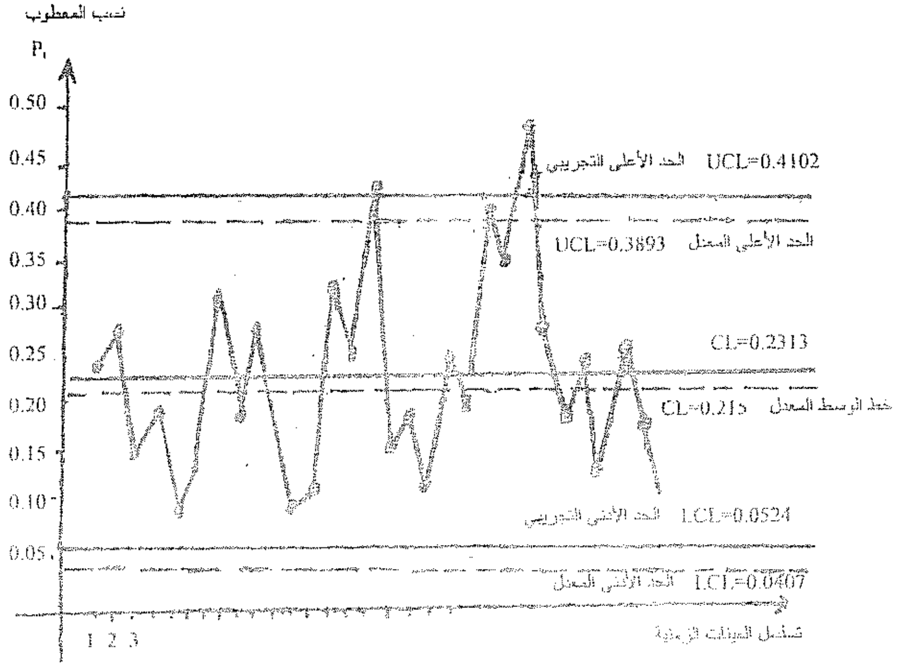
$$\overline{p} = \frac{\sum_{i=1}^{30} \hat{p}_i}{30} = 0.2313 = CL$$

وبالاستفادة من العلاقات (9.29)،(9.28) نجد أنّ:

$$UCL_p = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0.2313 + 0.1789 = 0.4102$$

$$LCL_p = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0.2313 - 0.1789 = 0.0524$$

وبهذه المعلومات وجدول البيانات السابق يمكننا تكوين خارطة الضبط لنسب الإنتاج المعيب كالتالي:



الشكل (14-9): يوضح خارطة ضبط نسب المعطوب من علب التعبئة لكل عينة من العينات العشوائية المأخوذة من المصنع وفقاً لجدول البيانات السابق.

يلاحظ من خارطة الضبط السابقة، شكل (14-9)، بأنّ هنالك نسبتين خارجتان عن حد الضبط الأعلى التجريبي وعليه نستنتج بأنّ العملية الإنتاجية ليست ضمن السيطرة أو غير محكمة ويتوجب دراسة هاتين العينتين (العينة 15 والعينة 23) لتحديد أسباب هذا الخلل واتخاذ الإجراء المناسب. وعلاوة على ذلك نجد أنّ انتشار أو تشتت نسب المعطوب عن خط الوسط CL كبيراً على الرغم من أنّ العملية الإنتادية ضمن حدود الضبط تقريباً وهو ما ينذر بخطورة تسريب العلب للحليب وبدراسة كل من العينتين 15، 23 وجد أنّ سبب الخروج الأول عن الانضباط هو تغيير نوعية العلب خلال النصف ساعة الخامسة عشرة بينما وجد أنّ سبب الخروج الثاني عن الانضباط هو تعيين عامل جديد لآلة التعليب.

وعلى ضوء هذه الأسباب غير الميكانيكية نجد عدم ضرورة عمل صيانة للآلة وأنّ الحل يكمن فقط في حذف أو إهمال العينتين 15 ، 23 من خارطة الضبط حيث أنّ إهمالهما وغض النظر عنهما هو الحل الصائب لخارطة الضبط السابقة وعلى ضوء الإجراء المذكور يتوجب رسم حدود الضبط الجديدة بدون العينتين 15 ، 23 وعليه نجد أنّ خطوط الضبط المعدّلة (الجديدة) (revised control lines) لنسب المعطوب في المثال السابق تأخذ القيم التالية:

خط الوسط المعدّل:

$$\overline{p} = \frac{301}{(28)(50)} = 0.2150$$

الحد الأعلى المعدّل:

$$UCL = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0.2150 + 3\sqrt{\frac{0.2150(0.7850)}{50}} = 0.3893$$

الحد الأدنى المعدّل:

$$LCL = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0.2150 - 3\sqrt{\frac{0.2150(0.7850)}{50}} = 0.0407$$

وقد تمّ رسمها بشكل خطوط نقطية في نفس خارطة الضبط لنفس المثال (الشكل (14-9)).

ويلاحظ من الشكل الجديد أيضاً بأنّ نسبة المعطوب في العينة رقم 21 أصبح خارج خط الانضباط العلوي المنقط وبدراسة البيانات مرة أخرى والعينة أيضاً لم نجد سبباً منطقياً أو مقبولاً وغير ميكانيكي لهذا الخروج عن خط الضبط العلوي وبالتالي لا يمكننا إهمال أو غض النظر عن هذه النسبة الخارجة عن الانضباط وعلى ضوء ذلك وبالإضافة إلى نوعية التشتت للنسب الذي يكون ضمن الوسط المعدّل CL نستنتج ضرورة صيانة آلة التعليب قيد الدراسة.

لاحظنا أنّ سبب ا لخروج الأول عن خط الانضباط العلوي لنسب العينتين 15 ، 23 كان إدارياً وليس ميكانيكياً وهو نوعية الورق وتعيين عامل جديد بدون خبرة بينما سبب الخروج الثاني عن خط الانضباط العلوي كان ميكانيكياً وليس إدارياً وبعد التعديل الإداري والصيانة الميكانيكية تم أخذ 24 عينة عشوائية حجم كل منها (50 منتجاً) وعلى فترات زمنية متساوية ($\frac{1}{2}$ ساعة خلال الدوريات الثلاث للآلة). وتم تسجيل أعداد المعيب في كل عينة مع حساب نسبها في الجدول (4-9) التالي:

جدول (4-9)

المجموعة الثانية للعينات: نسب المعطوب في العينات المأخوذة بعد التعديل والصيانة

تسلسل العينات i	عدد المنتجات المعطوبة d_i	النسب للمنتجات المعطوبة $\hat{p}_i$
1	9	0.18
2	6	0.12
3	12	0.24
4	5	0.10
5	6	0.12
6	4	0.08
7	5	0.10
8	3	0.06
9	7	0.14
10	6	0.12
11	2	0.04
12	4	0.08
13	3	0.06
14	6	0.12
15	5	0.10
16	4	0.08
17	8	0.16
18	5	0.10
19	6	0.12
20	7	0.14
21	5	0.10
22	6	0.12
23	3	0.06
24	4	0.08
	المجموع = 133	المتوسط $\overline{p}_2 = 0.1108$

ويمكننا هنا من الناحية النظرية وقبل تكوين خارطة أو مخطط الضبط لمجموعة العينات الثانية (بعد التعديل والصيانة) إجراء اختبار الفرضيات لمجموعتي العينات قبل وبعد التعديل والصيانة وهما:

1- فرضية العدم (أي عدم الاختلاف بينهما) $H_0 : \bar{p}_1 = \bar{p}_2$

2- فرضية البديل (الاختلاف بين المجموعتين) $H_1 : \bar{p}_1 > \bar{p}_2$

(أو أنّ متوسط النسب لأحدهما أكبر من الثاني على الأقل)

حيث وجدنا أنّ متوسط النسبة في مجموعة العينات الأولى قبل التعديل والصيانة هو $\bar{p}_1 = 0.215$ بينما نجد أنّ متوسط النسب بعد التعديل والصيانة في مجموعة العينات الثانية هو :

$$\bar{p}_2 = \frac{\sum_{i=1}^{m} d_i}{(50)(24)} = \frac{133}{1200} = 0.1108$$

أمّا الدرجة المعيارية للاختبار [راجع اختبار الفرضيات (الفصل السادس)]

$$z_0 = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

حيث تكون النسبة التجميعية للمنتجات المعطوبة:

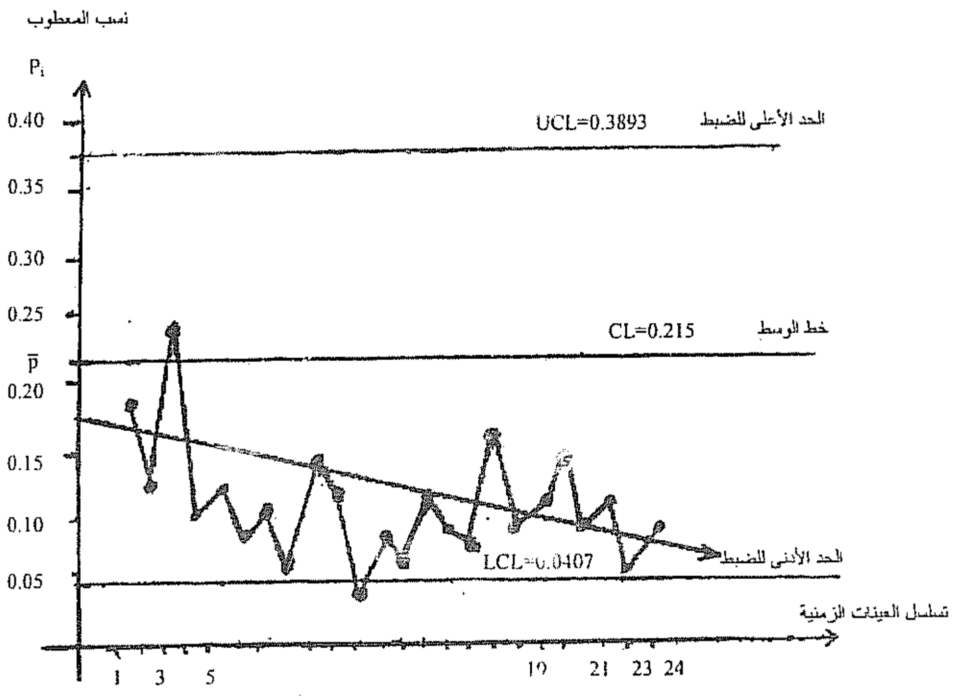
$$\hat{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{(1400)(0.215) + (1200)(0.1108)}{1400 + 1200} = 0.1669$$

$$z_0 = \frac{0.215 - 0.1108}{\sqrt{(0.1669)(0.8331)(\frac{1}{1400} + \frac{1}{1200})}} = 7.1078$$

وبالمقارنة مع الدرجة المعيارية الحرجة z_α لمنطقة الرفض $\alpha = 0.05$ المتعارفة نجد أنّ:

$$z_0 = 7.1078 > z_{\alpha = 0.05} = 1.645$$

وعليه نرفض فرض التساوي H_0 ونقبل الفرض البديل H_1 لأنّ z_0 تقع داخل منطقة الرفض أي نستنتج بأنّ هنالك نقصاناً جوهرياً لا يمكن غض النظر عنه في نسب منتجات المجموعة الثانية كما يوضحه مخطط الضبط للمجموعة الثانية (الشكل (9-15)) والمحسوبة بنفس الأسلوب السابق للمجموعة الأولى تالياً:



الشكل (9-15): مخطط الضبط لنسب المعطوب p_i في مجموعة العينات الثانية بعد الصيانة

ويلاحظ من مخطط الضبط أعلاه، شكل (9-15)، بأنّ هنالك انحداراً حاداً في خط التوجه العام لسلسلة النسب أي أنّ هنالك نقصاناً متسارعاً في نسب المنتجات المعطوبة p_i لكل عينة عن خط الوسط والاعتدال مما ينذر بخطر خروج العملية عن خط الضبط الأدنى علماً بأنّ نسبة المعطوب (التالف) في العينة 11 من المجموعة الثانية هي خارج حد الضبط الأدنى. ويبدو من خارطتي الضبط للمجموعتين الأولى والثانية، الشكل (9-14) والشكل (9-15) ضرورة إعادة النظر بحدود الضبط وأخذ مجموعة ثالثة من العينات تتكون من 40 عينة متتالية كل نصف ساعة وعلى مدى الورديات الثلاث ويحتوي كل منها على 50 عنصراً حيث تمّ تسجيل كل من أعداد المنتجات غير المتطابقة مع المواصفات (المعطوبة) ونسبها في الجدول (9-5) التالي:

جدول (5-9) المجموعة الثالثة من العينات للمثال (4) بعد التعديل والصيانة

تسلسل العينات i	أعداد المعطوب d_i	نسب المنتجات المعطوبة $\hat{p}_i = \frac{d_i}{n}$
1	8	0.16
2	7	0.14
3	5	0.10
4	6	0.12
5	4	0.08
6	5	0.10
7	2	0.04
8	3	0.06
9	4	0.08
10	7	0.14
11	6	0.12
12	5	0.10
13	5	0.10
14	3	0.06
15	7	0.14
16	9	0.18
17	6	0.12
18	10	0.20
19	4	0.08
20	3	0.06
21	5	0.10
22	8	0.16
23	11	0.22
24	9	0.18
25	7	0.14
26	3	0.06
27	5	0.10
28	2	0.04
29	1	0.02
30	4	0.08
31	5	0.10
32	3	0.06
33	7	0.14
34	6	0.12
35	4	0.08
36	4	0.08
37	6	0.12
38	8	0.16
39	5	0.10
40	6	0.16
	المجموع = 218	المتوسط $\bar{p}_3 = 0.1108$

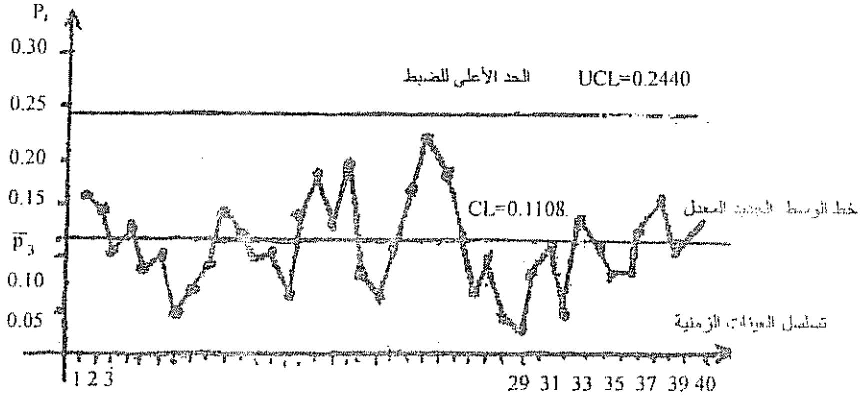
ومنه نجد أنّ حدود الضبط الجديدة المعدّلة هي كالتالي وكما في مخطط الضبط المبيّن بالشكل (9-16):

$$CL = \bar{p}_3 = 0.1108 = \bar{p}_3$$

$$UCL = \bar{p}_3 + 3\sqrt{\frac{\bar{p}_3(1-\bar{p}_3)}{n}} = 0.1108 + 3\sqrt{\frac{(0.1108)(0.8892)}{50}} = 0.2440$$

$$LCL = \bar{p}_3 - 3\sqrt{\frac{\bar{p}_3(1-\bar{p}_3)}{n}} = 0.1108 - 3\sqrt{\frac{(0.1108)(0.8892)}{50}} = -0.0224$$

ونظراً لكون خط الضبط الأدنى أصغر من الصفر $LCL < 0$ لذا نهمل أو نغض النظر عن القيمة السالبة ونستعيض عنها بخط الضبط الصفري أي اعتماد $LCL = 0$. وبهذا يكون لخريطة أو مخطط الضبط الجديد للعملية الإنتاجية حداً علوياً فقط وكما في المخطط المبيّن في الشكل (9-16) التالي:



الشكل (9-16): مخطط (خارطة) الضبط للمجموعة الثالثة من العينات في المثال رقم (4) لنسب المعطوب من علب التعبئة.

ويلاحظ من المخطط (9-16) بأنّ العملية الإنتاجية منضبطة (محكمة) للمرحلة الحالية كما نلاحظ أنّ نسب المعطوب للعينات تتقارب مع خط الوسط أو خط الاعتدال على الرغم من أنّ نسبة التسريب (أو التالف أو المعطوب من العلب) عالية أيضاً – بخلاف المطلوب وهو عدم التسريب مطلقاً – ولكنها تعتبر مقبولة.

إنّ خرائط نسب المعطوب $p_i = \frac{d_i}{n}$ من الإنتاج في العينة الواحدة تعتمد على ثلاثة عناصر:

1- حجم العينة 2- زمن وترتيب أخذ العينة 3- بعد حدّي الضبط الأعلى والأدنى عن بعضهما أو لنقل تشتت النسب عن خط الوسط أو خط الاعتدال المطلوب وفي حالة خرائط النسب p $(p - charts)$ يتوجب أن يكون حجم العينة كبيراً لضمان الحصول على احتمال أكبر في اكتشاف وقوع منتجات معطوبة في العينة وبما أننا نرغب في عدم الحصول على حدٍ أدنى سالب لضبط الجودة أي نرغب في أن يكون:

$$LCL = p - k\sqrt{\frac{p(1-p)}{n}} > 0 \qquad \textbf{(9.30)}$$

لذا نجد أنّ:

$$n > \frac{1-p}{p}k^2 \qquad \textbf{(9.31)}$$

حيث k يمثل عدد وحدات الانحراف المعياري σ للمجتمع الإحصائي ذي التوزيع الثنائي ولو افترضنا أنّ $p = 0.05$ مثلاً، وأنّ حد الضبط المستخدم هو $k\sigma = 3\sigma$ حسب المتعارف فإنّ حجم العينة في هذه الحالة يتوجب أن يكون:

$$n > \frac{0.95}{0.05}(3^2) = 171$$

ومن هنا نلاحظ بأنّ نظرية الحد الأدنى الموجب تفيد في حساب قانون حجم العينة في خرائط النسب.

ملاحظة ①: بما أننا نتعامل مع نسب المنتجات المعطوبة في العينة الواحدة $p_i = \frac{d_i}{n}$ يمكننا استخدام أحجام مختلفة (n_i) للعينات بدلاً من الحجم الواحد الثابت n أي أنّ p_i تكون بالشكل $p_i = \frac{d_i}{n_i}$ وفي مثل هذه الحالة المتغيرة الأحجام n_i سنحصل على حدود متغيرة أيضاً لضبط الجودة والمراقبة لكل عينة من العينات وبهذا سنحصل على حد أعلى وأدنى للضبط متدرج أو متموج وغير مستقيم حيث نجد أنّ:

$$CL = \overline{p} = \frac{\sum_{i=1}^{m} d_i}{\sum_{i=1}^{m} n_i} \qquad \text{خط الوسط:} \qquad \textbf{(9.32)}$$

حيث m عدد العينات المأخوذة. أمّا حدي الضبط العلوي والسفلي فيأخذان الصيغة التالية:

$$UCL_i = \bar{p} + 3\sigma_{\hat{p}} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}} \qquad (9.33)$$

$$LCL_i = \bar{p} - 3\sigma_{\hat{p}} = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}} \qquad (9.34)$$

وهنالك تعديل على الطريقة السابقة وهو استخدام متوسط الأحجام المختلفة للعينات $\bar{n} = \frac{\sum_{i=1}^{m} n_i}{m}$ حيث m عدد العينات، بدلاً من n_i في كل من العلاقتين السابقتين (9.33) و (9.34). لنحصل على علاقات جديدة لحدود الضبط العليا والدنيا للنسب.

9-10 خرائط أعداد العيوب (c – charts):

يرمز الحرف c لأعداد العيوب أو عدم التطابق مع المواصفات المطلوبة (non conformities) في منتجات العينة الواحدة المأخوذة من العملية الإنتاجية وعادةً ما تؤخذ أحجام العينات في مثل هذه الحالة صغيرة $n \geq 1$ مثلاً فإذا كان عدد العيوب في العينة الصغيرة كبيراً فإنّ ذلك مؤشر ينذر بالخطر وعليه يمكن تكوين نوعين من خرائط C يعتمد أحدهما على عدد العيوب في الوحدة الإنتاجية الواحدة. بينما يعتمد النوع الثاني على متوسط العيوب $\bar{c}$ في العينة الواحدة (أي عدد العيوب للوحدة الواحدة)، علاوة على افتراض ثبوت حجم العينة بحيث تتناسب مع توزيع بواسون حيث يتساوى المتوسط مع التباين وهذا يعني أنّ عدد فرص حصول عدم التطابق مع المواصفات (العيوب) كبير جداً وقد يؤول إلى ∞ بينما احتمال وقوع أو حصول أي عيب في أي موقع أو مرحلة من المراحل صغير جداً وثابت وعلى هذا الأساس نفترض أنّ:

$$x = 0, 1, 2, \ldots$$

هو المتغير العشوائي الذي يمثل عدد العيوب في الوحدة الإنتاجية الواحدة أو مجموعة من الوحدات الإنتاجية وأنّ العدد c يمثل كلاً من المتوسط μ والتباين σ^2 المعلومين في هذا التوزيع (توزيع بواسون μ=c=σ^2) حيث:

$$P(x) = \frac{e^{-c} c^x}{x!}$$

ومنه يمكن تكوين حدود الضبط كالتالي:

$$CL = c$$

$$UCL = c + 3\sqrt{c} \qquad (9.35)$$

$$LCL = c - 3\sqrt{c} \qquad (9.36)$$

وعندما يكون $LCL < 0$ سالباً فإننا نستعيض عنه بالصفر $(LcL = 0)$. وفي حالة عدم معرفة كل من المتوسط μ والتباين σ^2 للمجتمع الإحصائي (الإنتاج قيد الدراسة) أي عند عدم معرفة c نستطيع استخدام التقدير النقطي للعدد c وذلك باستخدام $\overline{c}$ (متوسط العيوب في العينات) أي أنّ حدود الضبط في مثل هذه الحالات ستكون كالتالي:

$$CL = \overline{c}$$

$$UCL = \overline{c} + 3\sqrt{\overline{c}} \qquad (9.37)$$

$$LCL = \overline{c} - 3\sqrt{\overline{c}} \qquad (9.38)$$

وتعتبر خرائط عدد العيوب c أكثر وضوحاً وموثوقية من خرائط النسبة p نظراً لتعددية أنواع وأسباب العيوب وتشخيصها.

مثال (5): لوحظت عيوب الإنتاج لأحدى الشركات المنتجة للدارات الكهربائية الالكترونية من خلال 26 عينة متتالية زمنياً تحتوي كل منها على مائة وحدة إنتاجية وسجلت أعداد العيــوب في الجدول التالي:

الجدول (9-6)

تسلسل العينات	عدد العيوب	تسلسل العينات	عدد العيوب
1	21	14	19
2	24	15	10
3	16	16	17
4	12	17	13
5	15	18	22
6	5	19	18
7	28	20	39
8	20	21	30
9	31	22	24
10	25	23	16
11	20	24	19
12	24	25	17
13	16	26	15

حيث يكون مجموع العيوب = 516 عيباً

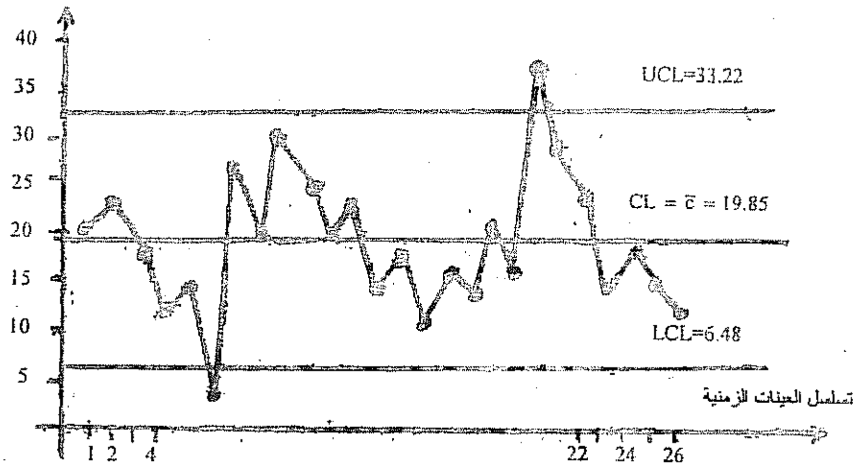
ومنها نجد أنّ حدود الضبط التجريبية (Trial cont. limits) لخارطة c تحسب باستخدام التقدير المتوسط $\bar{c}$ نظراً لعدم معلوميه كل من μ، σ^2 أي أنّ:

$$CL = \bar{c} = \frac{516}{26} = 19.85$$

$$UCL = \bar{c} + 3\sqrt{\bar{c}} = 19.85 + 3\sqrt{19.85} = 33.22$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}} = 19.85 - 3\sqrt{19.85} = 6.48$$

وبهذا تكوّن خارطة c لأعداد العيوب كما في الشكل التالي:



الشكل (17-9): خارطة c لأعداد العيوب في المثال (5)

ويلاحظ من الخارطة السابقة (الشكل (17-9)) بأنّ هنالك نقطتين خارج حدود الضبط التجريبية تمثلان أعداد العيوب في كل من العينتين 6 ، 20 وبعد دراسة كل من العينتين وجد أنّ الخلل إداري ناتج عن تبديل مراقب خط الإنتاج وعدم معرفته برصد العيوب في العينة 6 بينما كان الخلل في الحالة الثانية حراري مما أدى إلى خلل في اللحام الموصل للكهربائية وبمعرفة هذه الأسباب وتصليح الخلل نستطيع إهمال أو غض النظر عن كل من العينتين 6 ، 20 وتكوين خارطة c الجديدة وبحدود الضبط المعدّلة كالتالي:

(حيث أصبح عدد العينات 24 وأصبح مجموع العيوب هو 472 بعد حذف العينتين 6 ، 20)

خط الوسط المعدّل $CL = \bar{c} = \frac{472}{24} = 19.67$

حدي الضبط العلوي والسفلي المعدّلين هما:

$$UCL = \bar{c} + 3\sqrt{\bar{c}} = 19.67 + 3\sqrt{19.67} = 32.97$$

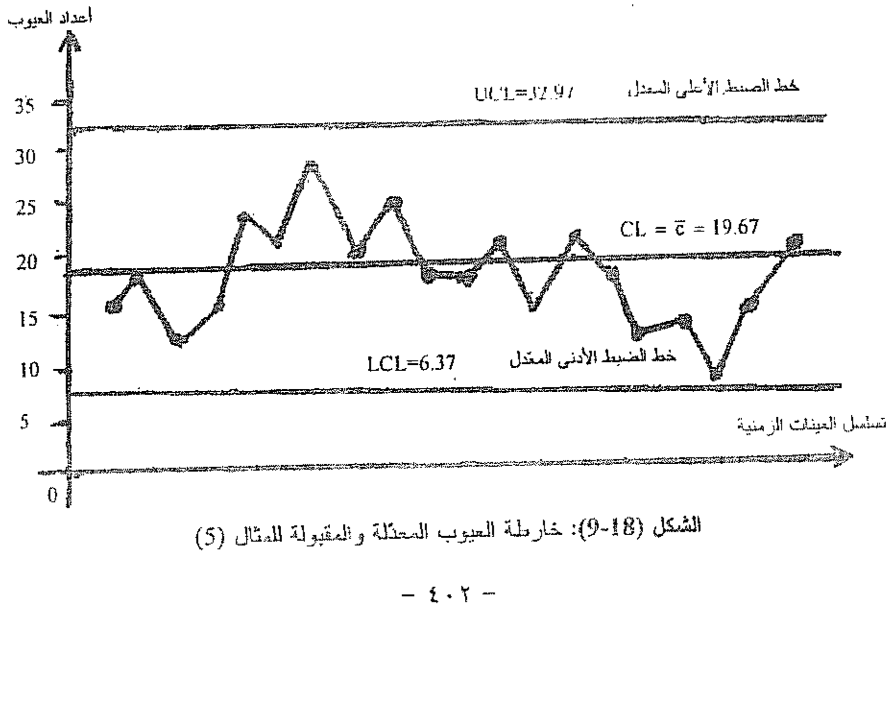
$$LCL = \bar{c} - 3\sqrt{\bar{c}} = 19.67 - 3\sqrt{19.67} = 6.37$$

وتعتبر هذه الحدود المعدلة مقياساً معتمداً لضبط المنتجات الجديدة بعد مرحلة التصليح والصيانة.

وبأخذ مجموعة جديدة من العينات بعد التصحيح متكونة من 20 عينة وبصورة متتالية، تمّ رصد العيوب فيها كما في الجدول التالي:

الجدول (7-9)

تسلسل العينة	عدد العيوب	تسلسل العينة	عدد العيوب	تسلسل العينة	عدد العيوب
1	16	8	20	15	19
2	18	9	25	16	12
3	12	10	19	17	14
4	15	11	18	18	9
5	24	12	21	19	20
6	21	13	16	20	21
7	28	14	22		



الشكل (18-9): خارطة العيوب المعدّلة والمقبولة للمثال (5)

وبمطالعة خارطة العيوب شكل (9-18) للمجموعة الجديدة نلاحظ عدم وجـود أي نقطـة أو قيمة $\bar{c}$ خارج حدود الضبط على الرغم من تشتتها الملحوظ حول الوسط وبهذا نستنتج بأنّ العملية الإنتاجية منضبطة أو محكمة.

9-11 خرائط أعداد العيوب u للوحدات الإنتاجية (u-charts):

يعتمد تكوين خرائط عدد عيوب الوحدات الإنتاجية $(u-\text{charts})$ على متوسط عدد العيوب للوحدة الإنتاجية الواحدة في العينة فإذا كان مجموع عدد العيوب هو c في عينة ذات حجم ثابت n فإنّ متوسط العيوب للوحدة الواحدة: $u=\frac{c}{n}$ حيث U عبارة عن متغير عشوائيّ لتوزيع بواسون أيضاً وحيث يعتبر مجموعاً خطياً لـ n من متغيرات بواسون العشوائية وعليه نجد أنّ حدود الضبط والمراقبة في هذه الخرائط تأخذ القيم التالية:

$$CL=\bar{u}=\frac{\sum_{i=1}^{m}u_i}{m} \qquad \text{خط الوسط:} \qquad (9.39)$$

حيث m عدد العينات العشوائية، $\bar{u}$ متوسط متوسطات العيوب للوحدة الواحدة.

$$UCL=\bar{u}+3\sqrt{\bar{u}/n} \qquad \text{الحد الأعلى للضبط:} \qquad (9.40)$$

$$LCL=\bar{u}-3\sqrt{\bar{u}/n} \qquad \text{الحد الأدنى للضبط:} \qquad (9.41)$$

مثال (6): يرغب أحد مهندسي الضبط والمراقبة لشركة خاصة لتصنيع أجهزة الكمبيوتر في تكوين خارطة ضبط لمراقبة عدد العيوب للوحدة الإنتاجية الواحدة. أخذت 20 عينة عشوائية متتالية ذات حجم ثابت $n=5$ وسجلت أعداد العيوب من هذه العينات كما في الجدول التالي:

الجدول (8-9)

التسلسل i	الحجم n	c_i مجموع العيوب	$u_i = \frac{c_i}{n}$ متوسط العيوب للوحدة	التسلسل i	الحجم n	c_i مجموع العيوب	$u_i = \frac{c_i}{n}$ متوسط العيوب للوحدة
1	5	10	2.0	11	5	9	1.8
2	5	12	2.4	12	5	5	1.0
3	5	8	1.6	13	5	7	1.4
4	5	14	2.8	14	5	11	2.2
5	5	10	2.0	15	5	12	2.4
6	5	16	3.2	16	5	6	1.2
7	5	11	2.2	17	5	8	1.6
8	5	7	1.4	18	5	10	2.0
9	5	10	2.0	19	5	7	1.4
10	5	15	3.0	20	5	5	1.0

حيث نجد أنّ مجموع العيوب = 193 ومجموع متوسطات العيوب للوحدات = 38.6

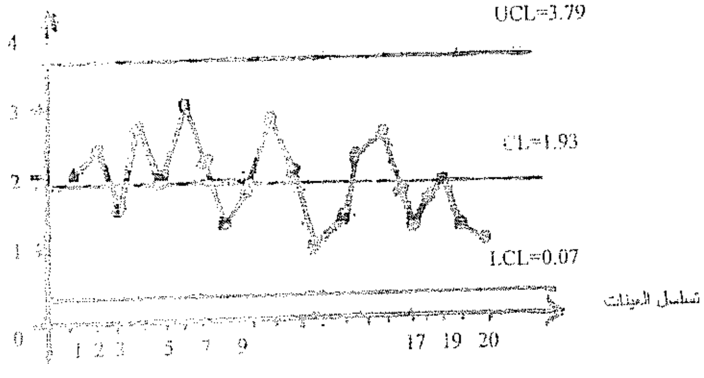
ومن المعلومات أعلاه نجد أنّ:

خط الوسط:
$$CL = \bar{u} = \frac{\sum_{i=1}^{20} u_i}{20} = \frac{38.60}{20} = 1.93$$

وأنّ: الحد الأعلى:
$$UCL = \bar{u} + 3\sqrt{\bar{u}/n} = 1.93 + 3\sqrt{1.93/5} = 3.79$$

الحد الأدنى:
$$LCL = \bar{u} - 3\sqrt{\bar{u}/n} = 1.93 - 3\sqrt{1.93/5} = 0.07$$

حيث نحصل على الخارطة التالية:



الشكل (19-9): خارطة العيوب للوحدة الواحدة المثال (6)

في كثير من حالات تعداد العيوب نصطدم بالواقع ونلاحظ عدم إمكانية توحيد أحجام العينات كما في المثال السابق. فمثلاً في مراقبة لشركات النسيج أو تصنيع الورق نجد عدم إمكانية توحيد أحجام العينات المأخوذة لأنّ لفات النسيج أو لفات الورق تختلف في أطوالها وفي عرضها فعند اختيارنا عشر لفات من النسيج نجد أنّ مساحتها الكلية تختلف عن مساحة عشر لفات أخرى مختارة عشوائياً من المصنع وكذلك بالنسبة للورق وفي مثل هذه الحالات نستخدم أسلوب الأحجام المتغيرة وغير الثابتة وعلى ضوء استخدام هذا الأسلوب نجد أنّ حدود الضبط العليا والدنيا تتغير هي الأخرى وليست ثابتة أحياناً أو متغيرة لبعض العينات وثابتة لمجموعة أخرى منها.

مثال (7): في مراقبة لإحدى شركات النسيج تم أخذ عشرة لفات متتالية من النسيج وتم رصد عدد العيوب على كل 50 متر مربع فيها وعلى ضوء المساحة الكلية لكل لفة أو ربطة من الوحدات المنتجة، كما في الجدول التالي لتكوين وبناء خارطة العيوب في الوحدة الواحدة أو اللفة الواحدة (u – charts).

الجدول (9-9)

تسلسل لفات النسيج i	المساحة الكلية بالمتر المربع	مجموع العيوب	تعداد المساحات في اللفة الواحدة	متوسط العيوب الوحدة الواحدة u_i
1	500	14	10	1.4
2	400	12	8	1.5
3	650	20	13	1.54
4	500	11	10	1.10
5	475	7	9.5	0.74
6	500	10	10	1.00
7	600	21	12	1.75
8	525	16	10.5	1.52
9	600	19	12	1.58
10	625	23	12.5	1.84
		مجموع العيوب الكلي: 153	مجموع وحدات المساحات المفحوصة: 107.50 وحدة	

ومن الجدول السابق نجد أنّ:

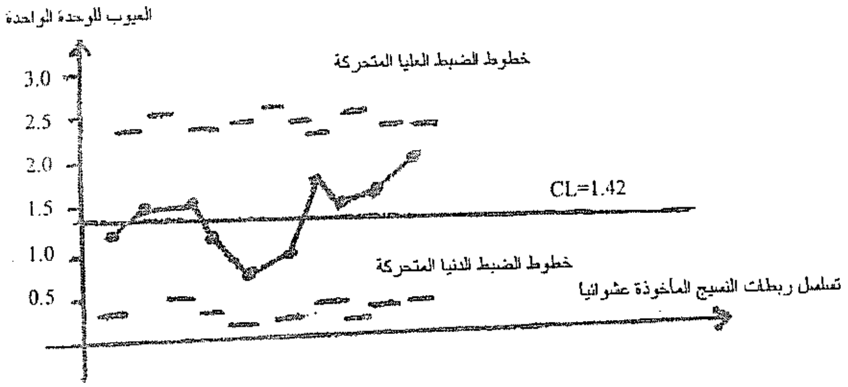
خط الوسط الثابت $CL = \bar{u} = \frac{153}{107.5} = 1.42$

وعلى ضوء تغير أعداد المساحات المفحوصة أو المراقبة في كل لفة من النسيج يتوجب استخدام تلك الأعداد n_i بدلاً من n (عدد المفردات في العينة) لكل لفة من النسيج بصورة منفردة لتكوين خطوط الضبط والمراقبة العليا والدنيا للبيانات السابقة وكما يلي:

الجدول (9-10)

تسلسل لفات النسيج	أعداد المساحات (50 م²) في كل لفة n_i	$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n_i}}$	$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n_i}}$
1	10	2.55	0.29
2	8	2.68	0.16
3	13	2.41	0.43
4	10	2.55	0.29
5	9.5	2.58	0.26
6	10	2.55	0.29
7	12	2.45	0.39
8	10.5	2.52	0.32
9	12	2.45	0.39
10	12.5	2.43	0.41

حيث نحصل على الشكل التالي:



الشكل (9-20): خارطة العيوب للوحدة الواحدة لأحجام متغيرة

9-12 توزيعات باريتو (pareto distribution):

وهي عبارة عن أشكال إحصائية وصفية تكرارية لعيوب الإنتاج وتعطى على أشكال مختلفة كمضلعات تكرارية أو نقطية حيث يتم فيها تحديد نوعية العيوب وعدم المطابقة مع المواصفات وتكراراتها أو نوعية العيوب ونسب تلك التكرارات وتعطى على شكل التكرارات المجتمعة الصاعدة عادة. وبديهي أنّ تشخيص نوعية العيوب يسهّل من عمليات الصيانة وتطوير الإنتاج سواءً كانت هذه العيوب عددية أو وصفية وعلى أساس أن الكثير من الأعطال هي صفة للنادر أو القليل من أنواع هذه الأعطال، يتم توزيع باريتو كما في الشكل التالي حيث نبدأ بالتكرارات الكبيرة عادة للأعطال وهكذا بصورة تنازلية:

الجدول (11-9)

التوزيع التكراري النسبي المتجمع الصاعد لأنواع الأعطال في عملية تصنيع قطـع الـدارات الكهربائية لباريتو

التسلسل	رقم أو نوع الخلل		التكرار	التكرار المتجمع	التكرار النسبي	التكرار النسبي المتجمع
1	لحام غير كافي	**	40	40	40.82	40.82
2	لحام بارد	********************	20	60	20.41	61.22
3	لحـــام مفتوح	*******	7	67	7.14	68.37
4	قطع غير مضبوطة	******	6	73	6.12	74.49
5	لحام	*****	5	78	5.10	79.59
6	مؤشـــر الاختبار	***	3	81	3.06	82.65
7	....	***	3	84	3.06	85.71
8	....	***	3	87	3.06	88.78
9	زيـــادة قطع	**	2	89	2.04	90.82
10	قطـــع معطوبة	**	2	91	2.04	92.86
11	قطـــع مفقودة	**	2	93	2.04	94.90
12	أســلاك غيـــر مضبوطة	*	1	94	1.02	95.92
13	خـــتم العامل	*	1	95	1.02	96.94
14	أختـــام مفقودة	*	1	96	1.02	97.96
15	لحـــام قصير	*	1	97	1.02	98.98
16	... تالفة	*	1	98	1.02	100.0

ويلاحظ من جدول باريتو السابق بأنّ الخلل الكبير هو أن اللحام في هذه الدارات غير كـافٍ وعليه يتوجب التركيز على كفايتها إمّا بالصيانة أو باستبدال العامل ...إلخ أمّا الخلـل الكبيـر الثاني فهو في اللحام البارد وعليه يتوجب الصيانة واستبدال العامل أو المواد الأولية ...إلـخ وبنفس الأسلوب يتم صيانة الأعطال واحداً بعد الآخر وعلى الترتيب النسبي لضمان الجودة.

وتعتبر طريقة باريتو لرصد ومراقبة العيوب من الطرق التقنية الهامة التي تعتمد ليس علـى تعداد العيوب أو نسبها بل على نوعيتها أيضاً لتحديد أولوية الصيانة وضمان الجودة.

الجدول (12-9)

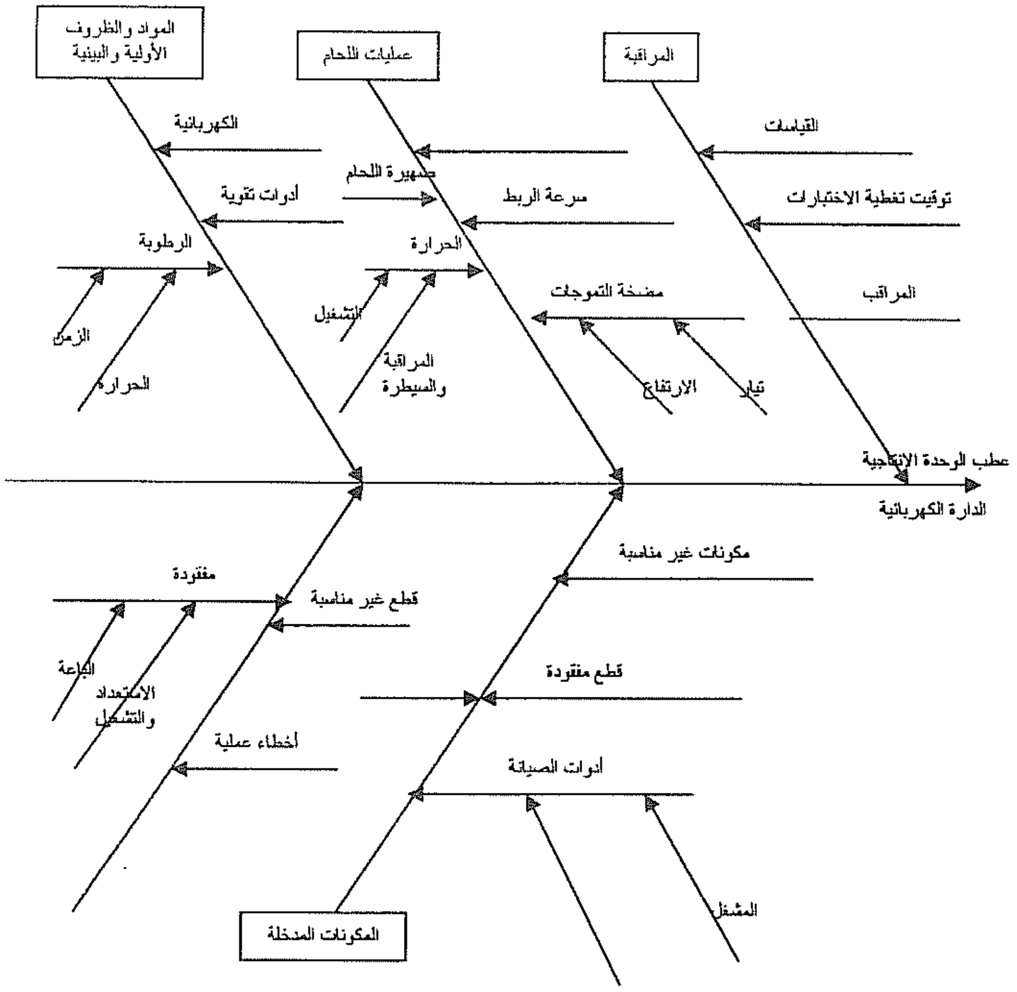
جدول العيوب المصنفة بأرقام القطع وتكراراتها ونسبهاالمئوية ورموز الأعطال لباريتو

رقم القطعة التكرار الكلي نسب السطر نسب العمود	[illegible]	[illegible]	[illegible]	رمز ثقب	رمز قصف	رمز قصف ثقب	لحام ناقص	لحام مفتوح DEWE	لحام بارد
0001285	1 -1.02 1.41 50.00	0 0.00 0.00 0.00	0 0 00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 1.02 1.41 100.00	0 0.00 0.00 0.00	5 5.10 7.04 71.43	20 20.41 28.17 100.00
0001481	1 1.02 3.70 50.00	2 2.04 7.41 100.00	2 6.12 22.22 100.00	6 3.06 11.11 100.00	3 0.00 0.00 100.00	0 0.00 0.00 0.00	1 1.02 3.70 100.00	2 2.04 7.41 28.57	0 0.00 0.00 0.00
0006429	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0 00 0.00	0 0.00 0.00 0.00
المجموع	2 2.04	2 2.04	2 2.04	6 6.12	3 3.06	1 1.02	1 1.02	7 7.14	20 20.41

رقم القطعة التكرار التكرار الكلي نسب السطر نسب العمود	[illegible]	[illegible]	[illegible]	[illegible] ID	[illegible] M	[illegible] EC	[illegible] 5	[illegible] (1)	المجموع
0001285	40 40.82 56.34 100.00	0 0 00 0.00 0.00	0 0.00 0 00 0.00	0 0.00 0.00 0.00	2 2.04 2.82 66.67	1 1.02 1.41 33.33	1 1.02 1.41 100.00	0 0.00 0.00 0.00	72.5
0001481	0 0.00 0.00 0.00	5 5.10 18.52 100.00	1 1.02 3.70 100.00	1 1.02 3.70 100.00	1 1.02 3.70 33.33	2 2.04 7.41 66.67	0 0.00 0.00 0.00	0 0.00 0.00 0.00	27.5
0006429	0 0.00 0.00 0.00	0 0 00 0.00 0.00	0 0.00 0 00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0 00	0.0
المجموع	40 40.82	5 5.10	1 1.02	1 1.02	3 3.06	3 3.06	1 1.02	0 0.00	0 100.00

9-13 مخططات الأسباب وتأثيراتها (Cause and effect diagram):

تعتبر مخططات الأسباب وتأثيراتها من التقنيات الهامة في تحليل عدم التطابق مع المواصفات المطلوبة للوحدات الإنتاجية وهي مخططات وصفية غير عددية تسمى بالهياكل السمكية أو أشكال إيشيكاوا نسبة إلى كورو إيشيكاوا. وتستخدم مخططات الأسباب وتأثيراتها لتوضيح أنواع الأعطال المختلفة لوحدات الإنتاج ومصادرها وعلاقاتها الداخلية فيما بينها، وتعتبر هذه المخططات كلوحة بيانات مفيدة لكل من العاملين والمهندسين الصناعيين ومدراء الإنتاج ومراقبي ومهندسي الجودة ومشاكلها. وبديهي أنّ جودة الهياكل السمكية هذه تعكس مستوى الفهم التقني للعملية الإنتاجية وأسرارها وهنالك العديد من الأساليب لرسم هذه المخططات أو الهياكل حيث ترسم هذه المخططات وفقاً لأولويات الجودة في العملية الإنتاجية كالتركيز مثلاً على المواد الأولية والعمال ووسائل الإنتاج وقطع الصيانة ...إلخ. وعادةً ما ترسم هذه المخططات وفقاً لتدفق المواد الأولية بدأ من اليسار إلى اليمين خلال العملية الإنتاجية وكما في المخطط التالي حول عملية إنتاج الدارات الكهربائية في المثال (5) السابق:



شكل (21-9): مخطط إيشيكاوا للأسباب وتأثيراتها للأعطال في الوحدات الإنتاجية

والحقيقة إنّ هذه المخططات تعتبر خلاصة إدارية هامة تستوحى عادةً من خلال جلسات كافة العاملين ضمن خط الإنتاج لدراسة وتحليل الأسباب والآثار الناتجة عن الأعطال في العملية الإنتاجية ومن ثم يتم اتخاذ القرار المناسب للصيانة والتطوير أو التحسين لكل مراحل العملية الإنتاجية بما فيها جودة الإنتاج.

9-14 قابلية أو قدرة العملية الإنتاجية (process capability):

تعتبر قابلية أو قدرة العملية الإنتاجية وحساباتها إحدى الوسائل الهامة أيضاً لتحسين الجودة وتطوير الإنتاج ونعني بالقابلية أو القدرة هو إمكانية العملية الإنتاجية على مطابقة المواصفات والشروط المفروضة والمحددة مسبقاً على العملية الإنتاجية (specification limits) وهي تختلف عن حدود الضبط المحسوبة سابقاً.

ولحساب ما إذا كانت العملية الإنتاجية على درجة من القدرة والقابلية لمطابقة المواصفات والشروط المفروضة مسبقاً يتم حساب ما يسمى بمعامل أو دليل القدرة (process capability index) والذي يُرمز له عادةً بالرمز PCI ويُعرَّف على أنه النسبة بين ما يسمى بمدى التحمل (Tolerance)، وهو الفرق بين الحدين الأعلى والأدنى للشروط المفروضة مسبقاً، إلى مدى الانحرافات المعيارية الستة للعملية الإنتاجية المستخدمة سابقاً في خرائط المراقبة أي الفرق بين الحدين الأعلى والأدنى المحسوبين في خرائط المراقبة وبعبارة رياضية يعرف معامل أو دليل القدرة PCI بالعلاقة التالية:

$$PCI = \frac{USL - LSL}{UCL - LCL} = \frac{USL - LSL}{6\sigma} \qquad (9.42)$$

ومن الواضح استقراء العلاقات التالية:

$PCI > 1$ يعني أنّ مدى تشتت المواصفات المشروطة أكبر وأوسع من مدى تشتت متوسطات العينات أو مقاييس الجودة عموماً في العملية الإنتاجية المنضبطة. وهذا يعني وقوع المنتجات ضمن المواصفات المشروطة وبمعنى آخر نستنتج بأنّ العملية الإنتاجية ذات قدرة أو قابلية تتطابق مع أو تفوق الشروط المفروضة للجودة.

$PCI = 1$ يعني تطابق أو تساوي مدى تشتت المواصفات مع مدى تشتت مقاييس الجودة في العملية الإنتاجية. وهذا يعني أيضاً أنّ العملية الإنتاجية ذات قدرة تتطابق تماماً ولا تفوق المواصفات والشروط المفروضة للجودة.

$PCI < 1$ يعني أنّ مدى تشتت المواصفات المشروطة أصغر أو أضيق من مدى تشتت مقاييس الجودة في العملية الإنتاجية المنضبطة. أي أن الحدين الأعلى والأدنى للضبط المحسوبين ضمن مخطط الضبط والمراقبة يقعان خارج المواصفات أو خارج الحدين الأعلى والأدنى المشروطين وبهذا سنجد بأنّ هنالك عدداً من هذه المنتجات خارج المواصفات المشروطة على الرغم من أنها داخل حدود الضبط المحسوبة للعملية الإنتاجية وحسب خارطة

الضبط التي تصف العملية الإنتاجية. كما تعني أيضاً بأنّ قدرة العملية الإنتاجية ليست ضمن الشروط المفروضة والمطلوبة أو بعبارة أخرى يُقال عن العملية الإنتاجية بأنها غير قادرة على مطابقة المواصفات المطلوبة ويتوجب عندئذ إجراء التعديل أو الصيانة والتطوير للمطابقة مع الشروط.

ففي المثال (1) لو افترضنا أن الشروط المفروضة على حدي الضبط الأعلى والأدنى لمتوسطات الإنتاج هو: $\mu \pm 3\sigma = 74 \pm 0.03$ أي أن: $LSL = 73.970$ وأنّ: $USL = 74.030$ بينما تشير حسابات المثال رقم (1) إلى أنّ حدي الضبط الأعلى والأدنى هما $LCL = 73.988 > LSL$ وأنّ: $UCL = 74.014 < USL$ أي أنّ هنالك اختلافاً بين حدي الضبط المفروضين وحدي الضبط المحسوبين أو الواقعيين (UNTL, LNTL) وأنّ هنالك نوعاً من الإزاحة بينهما مما يتوجب عليه حساب دليل أو معامل القابلية PCI لمعرفة كم من المتوسطات خارج شروط المطلوبة على الرغم من أنها داخل حدود الضبط المحسوبة أو الواقعية للعملية الإنتاجية وعليه نجد أنّ:

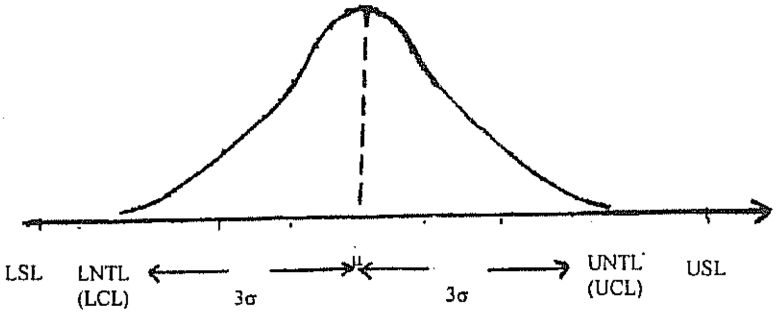
$$PCI = \frac{USL - LSL}{3\sigma - (-3\sigma)} = \frac{74.03 - 73.97}{6(0.0099)} = \frac{0.06}{0.0594} = 1.01$$

وهذا العدد PCI=1.01 يكاد يساوي الواحد مما يعني أنّ قدرة العملية الإنتاجية تكاد تتطابق مع المواصفات والشروط المفروضة أو أنّ مدى التحمل الواقعي للعملية الإنتاجية يكاد يتساوى مع الحدود والشروط المطلوبة أي أنّ هنالك عدداً ضئيلاً جداً ولا يعتد به سيكون خارج المواصفات والحدود أو الشروط المفروضة.

وللحصول على نسبة أفضل لأعداد أو نسب الوحدات المعطوبة والتالفة نستخدم الدرجة المعيارية على افتراض أنّ متوسطات حلقات المكابس للسيارات تتوزع توزيعاً طبيعياً $N(\mu, \sigma)$ وعليه نجد أنّ احتمالية العيوب أي احتمالية وقوع الوحدات المنتجة في منطقة الرفض على جهتي التوزيع هو احتمال وقوع متوسط المنتج دون الحد الأدنى LSL وفوق الحد الأعلى USL كما يلي:

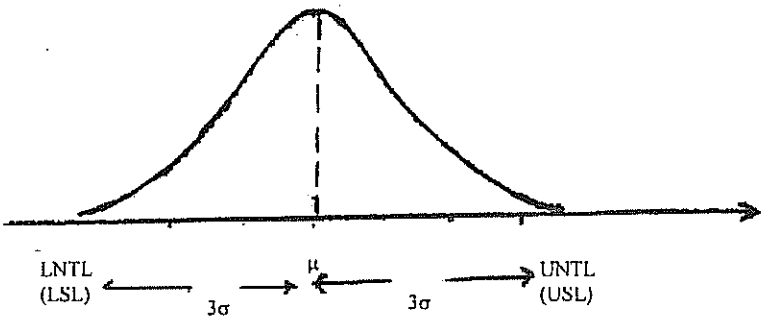
$$\begin{aligned}\hat{p} &= p(X < 73.970) + p(X > 74.03)\\ &= p\left(\frac{73.970 - 74.001}{0.0099}\right) + \left(1 - p\left(\frac{74.03 - 74.001}{0.0099}\right)\right)\\ &= p(-3.13) + 1 - p(2.93) = 0.00087 + 1 - 0.99831 = 0.00256\end{aligned}$$

ويعني هذا أنّ هنالك 256 وحدة منتجة معطوبة من بين كل مائة ألف وحدة منتجة أي أنّ نسبة حلقات المكابس غير المطابقة للمواصفات والشروط المطلوبة هـي 0.256٪ خـارج حـدي الضبط LSL ، USL. ويمكن توضيح ذلك بيانياً بالأشكال التالية:



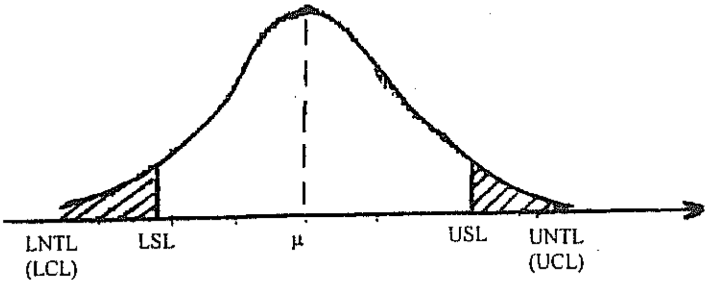
PCI > 1 -(a)

حدود الضبط المحسوبة ضمن المواصفات المشروطة والقدرة تفوق الشروط



PCI = 1 -(b)

تطابق بين الحدود الطبيعية المحسوبة للعملية الإنتاجية مع الحدود المشروطة والقدرة مطابقة للشروط والمواصفات



PCI < 1 -(c)

العملية الإنتاجية ذات قدرة غير مطابقة للشروط المفروضة وهنالك عدد من الوحدات الإنتاجية خارج الحدود المشروطة على الرغم من أنها داخل حدود الضبط الطبيعية

شكل (9-22)

ومن الجدير بالذكر بأنّ الدليل PCI يمكن أن يعرّف ويحسب لكل منطقة من مناطق خارطة الضبط الستة A ، B ، C حيث نلاحظ من خرائط الضبط بأنّ:

$$PCI_c = \frac{USL - LSL}{2\sigma}$$

$$PCI_B = \frac{USL - LSL}{4\sigma}$$

$$PCI_A = PCI = \frac{USL - LSL}{6\sigma}$$

مثال (8): استلمت شركة لتصنيع أقراص كمبيوتر (3.5 inch) شكاوى عن كثير من الأقراص لا تخزّن البيانات بصورة مضبوطة، وعليه قررت الإدارة تكوين خارطة لضبط العملية فأخذت عينة عشوائية كل ساعة، كل منها ذات حجم n = 200 قرص حيث يتم فحص كل قرص من الأقراص لمعرفة ما إذا كان معطوباً أم لا وسجلت النتائج بالجدول التالي:

جدول (13-9)

العينات	عدد الوحدات المعطوبة	العينات	عدد الوحدات المعطوبة	العينات	عدد الوحدات المعطوبة
1	19	14	10	27	9
2	5	15	18	28	10
3	16	16	20	29	10
4	20	17	13	30	18
5	6	18	6	31	15
6	12	19	8	32	16
7	18	20	3	33	5
8	6	21	8	34	14
9	13	22	7	35	3
10	15	23	4	36	10
11	10	24	19	37	19
12	6	25	3	38	13
13	7	26	19	39	19
				40	9

ولتكوين جدول نسبة المعطوب (p-chart) نحسب النسب أولاً لكل عينة من العينات:

حيث $p_i = \dfrac{d_i}{200}$

ومنها نحسب متوسط النسب: $\bar{p} = \dfrac{\sum_{i=1}^{40} p_i}{40} = 0.5762$

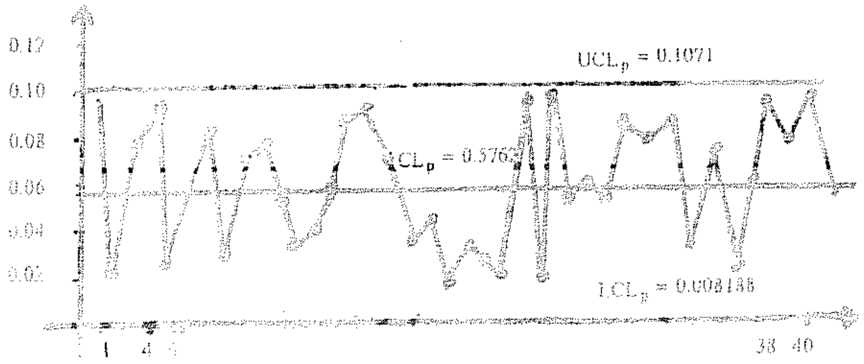
وعليه فإنّ حدي الضبط الأعلى والأدنى

$$UCL_p = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.5762 + 3\sqrt{\frac{0.5762(1-0.5762)}{200}} = 0.1071$$

$$LCL_p = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.5762 - 3\sqrt{\frac{0.5762(1-0.5762)}{200}} = 0.008188$$

وبنفس الأسلوب يمكن حساب حدود ضبط المناطق الأخرى الستة حيث نأخذ $k = \pm 1, \pm 2$

ونحصل على الشكل التالي:



شكل (9-23)

حيث نلاحظ بأن العملية الإنتاجية منضبطة أو محكمة على الرغم من تشتت النقاط حول CL

9-15 تمارين عامة:

1- البيانات التالية تمثل قيم $\bar{X}$ ، R لأربع وعشرين عينة عشوائية زمنية كل منهما ذات حجم $n = 5$ أخذت من مصنع المحامل الكروية حيث أخذت قياسات الأقطار الداخلية للمحامل:

رقم العينة	$\bar{x}$	R	رقم العينة	$\bar{x}$	R
1	0.50345	3	13	0.50354	8
2	0.50342	4	14	0.5034	6
3	0.50316	4	15	0.50371	5
4	0.50315	4	16	0.50349	7
5	0.50350	5	17	0.50335	4
6	0.50341	6	18	0.50317	3
7	0.50326	4	19	0.50340	8
8	0.50338	3	20	0.50351	4
9	0.50348	7	21	0.50337	2
10	0.50336	8	22	0.50328	1
11	0.50319	3	23	0.50335	3
12	0.50386	9	24	0.50342	2

والمطلوب:

أ – كون كُلاً من خارطتي $\bar{X}$ ، R لبيانات العملية الإنتاجية أعلاه وحدّد ما إذا كانت العملية ضمن حدود السيطرة والضبط وإن لم تكن فأعد صياغة حدود الضبط التجريبية.

ب– إذا كانت الحدود المشروطة للقطر هي أن يكون ضمن فترة التقدير 0.5030 ± 0.001 فأوجد كلاً من نسبة عدم التوافق ودليل القدرة PCI.

2- كوّن جدولي الضبط $\bar{X}$ ، R للبيانات التالية ووضح ما إذا كانت العملية الإنتاجية منضبطة أم لا.

رقم العينة	العناصر				رقم العينة	العناصر			
	x_1	x_2	x_3	x_4		x_1	x_2	x_3	x_4
1	6	9	10	15	11	8	12	14	16
2	10	4	6	11	12	6	13	9	11
3	7	8	10	5	13	16	9	13	15
4	8	9	6	13	14	7	13	10	12
5	9	10	7	13	15	11	7	10	16
6	12	11	10	10	16	15	10	11	14
7	16	10	8	9	17	9	8	12	10
8	7	5	10	4	18	15	7	10	11
9	9	7	8	12	19	8	6	9	12
10	15	16	10	13	20	14	15	12	16

3- تسحب عينة عشوائية منتظمة ذات حجم $n = 5$ كل نصف ساعة لضبط ومراقبة طول المنتجات الصناعية بالسنتمترات في إحدى العمليات الإنتاجية حيث سجلت القياسات بالجدول التالي:

تسلسل العينات ضمن الفترة الزمنية	البيانات				
	x_1	x_2	x_3	x_4	x_5
1	13	17	14	16	14
2	11	10	9	15	16
3	11	12	17	14	15
4	12	9	18	20	13
5	8	10	12	14	16
6	10	14	16	18	20
7	15	20	18	19	14
8	12	16	17	17	14
9	14	16	18	18	14
10	10	12	13	14	16

استخدم كلاً من خريطتي $\bar{X}$ ، R وحدد مناطقها الستة $(\pm\sigma, \pm2\sigma, \pm3\sigma)$ لمعرفة ما إذا كانت العملية منضبطة أم لا، ثم ناقش كلا الخارطتين وكوّن قراراً عن العملية الإنتاجية.

4- أخذت 50 عينة عشوائية بانتظام كل منها ذات حجم ثابت $n = 8$ من عملية إنتاجيـــة وتـــم حساب كل من قيم $\bar{x}$ ، R لكل عينة من العينات ووجد أنّ:

$$\sum_{i=1}^{50} R_i = 250 \qquad \sum_{i=1}^{50} \bar{x}_i = 2000$$

وعلى افتراض أن توزيع إحصائيات الجودة طبيعياً

أ – أحسب حدود الضبط لخرائط $\bar{X}$ ، R.

ب– إذا كانت العملية منضبطة في كلا الخارطتين فما هي حدود التحمـــل الطبيعيـــة للعمليـــة الإنتاجية.

جـ – إذا اشترطنا حدود الضبط USL ، LSL لتكون 41 ± 5.0 فما هي قابلية العملية الإنتاجية لإنتاج وحدات ضمن هذه الحدود المشروطة.

د – إذا افترضنا أنّ الوحدات المنتجة فوق حد الضبط العلوي المشروط USL يمكـــن إعـــادة تصنيعها بينما الوحدات تحت مستوى حد الضبط الأدنى المشروط LSL سوف تهمل، فما هي نسبة الوحدات المهملة والمعاد تصنيعها في العملية الإنتاجية.

هـ – أحسب معامل قابلية أو قدرة العملية الإنتاجية PCI واقترح كيفية التحسين.

5- أخذت 50 عينة عشوائية بانتظام، كل منها ذات حجم ثابت $n = 6$ وحســـبت إحصـــائيات الجودة $\bar{x}_i$ و S_i لكل عينة من العينات وسجلت كالتالي:

$$\sum_{i=1}^{50} S_i = 75 \quad , \quad \sum_{i=1}^{50} \bar{x}_i = 1000$$

فأحسب:

أ – خرائط الضبط لكل من $\bar{X}$ ، S على افتراض أنّ التوزيع طبيعي.

ب– إذا افترضنا أنّ العملية منضبطة فأحسب مدى التحمل للعملية.

جـ – إذا اشترطنا حدود الضبط USL, LSL لتكون 19 ± 4.0 فأحسب قابلية أو قدرة العمليـــة لإنتاج وحدات غير معيبة.

د – إذا كانت الوحدات المعيبة فوق حد الضبط USL قابلة لإعادة التصنيع بينما الوحدات دون حد الضبط LSL تهمل، فما هي نسبة عدد الوحدات المهملة والقابلة لإعادة التصنيع في هـــذه العملية.

6- إذا كانت حدود الضبط لكل من خارطتي $\overline{X}$، S كالتالي علماً بأنّ حجم العينات ثابت (n = 5) وأنّ كلا العمليتين منضبطتين (محكمتين):

خارطة S	خارطة $\overline{x}$
CL=1.000	CL=200
UCL=2.266	UCL=201.88
LCL=0	LCL=198.12

فما هي كفاءة العملية لإنتاج وحدات تتطابق مع المواصفات أو حدود الضبط LSL, USL المشروطة التالية 202.50 ، 197.50.

7- كوّن خارطة النسبة p لعمليات الإنتاج المستقبلية إذا علمت أنّ عشر عينات عشوائية وذات حجم n = 100 وحدة إنتاج تؤخذ على فترات منتظمة (عينة واحدة في كل يوم) حيث لوحظت النتائج التالية:

عدد الوحدات المعطوبة	رقم العينة	عدد الوحدات المعطوبة	رقم العينة
2	6	3	1
1	7	2	2
2	8	6	3
0	9	2	4
5	10	7	5

ثم أحسب حجم أصغر عينة يمكن استخدامها في مثل هذه العمليات الإنتاجية ليكون الحد الأدنى للضبط موجباً أي LSL > 0

8- لوحظت أعداد الأحزمة المعيبة في إحدى عمليات الإنتاج لصناعة الأحزمة المطاطية على ربطات ذات حجم 2300 حزاماً حيث تم رصد ومراقبة الربطات العشرين الأخيرة من العملية ورصدت النتائج والملاحظات التالية:

عدد الأحزمة المعيبة	رقم الربطة	عدد الأحزمة المعيبة	رقم الربطة
456	11	230	1
394	12	435	2
285	13	221	3
331	14	346	4
198	15	230	5
414	16	327	6
131	17	285	7
269	18	311	8
221	19	342	9
407	20	308	10

والمطلوب:

أحسب حدود الضبط التجريبية لخريطة النسبة p (نسب الوحدات المعيبة) واستخدم التعديل المطلوب لتكوين خارطة النسبة p لضبط عملية الإنتاج المستقبلية.

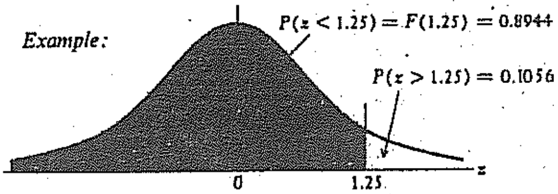
ملحق

الجداول الإحصائية

ملحق (1) – التوزيع الطبيعي المعياري

Normal distribution

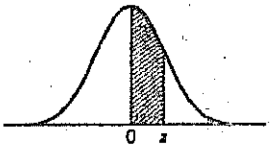
$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$ *Example:*



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

ملحق (2) – مساحات التوزيع المعتدل المعياري

المدخل للجدول هو النسبة من المنحنى كله ، والتي تنحصر بين 0 = z وقيمة z الموجبة . المساحات لقيم z السالبة نحصل عليها من التماثل .

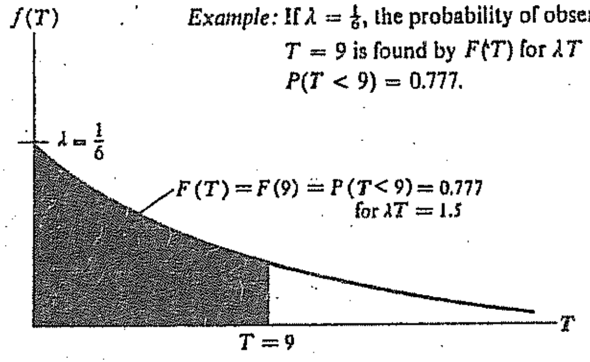


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

ملحق (3) - التوزيع الأسي

Exponential distribution

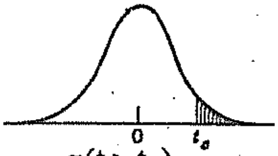
Example: If $\lambda = \frac{1}{6}$, the probability of observing a value less than $T = 9$ is found by $F(T)$ for $\lambda T = \frac{1}{6}(9) = 1.5$; $P(T < 9) = 0.777$.



λT	$F(T)$	λT	$F(T)$	λT	$F(T)$	λT	$F(T)$
0.0	0.000	2.5	0.918	5.0	0.9933	7.5	0.99945
0.1	0.095	2.6	0.926	5.1	0.9939	7.6	0.99950
0.2	0.181	2.7	0.933	5.2	0.9945	7.7	0.99955
0.3	0.259	2.8	0.939	5.3	0.9950	7.8	0.99959
0.4	0.330	2.9	0.945	5.4	0.9955	7.9	0.99963
0.5	0.393	3.0	0.950	5.5	0.9959	8.0	0.99966
0.6	0.451	3.1	0.955	5.6	0.9963	8.1	0.99970
0.7	0.503	3.2	0.959	5.7	0.9967	8.2	0.99972
0.8	0.551	3.3	0.963	5.8	0.9970	8.3	0.99975
0.9	0.593	3.4	0.967	5.9	0.9973	8.4	0.99978
1.0	0.632	3.5	0.970	6.0	0.9975	8.5	0.99980
1.1	0.667	3.6	0.973	6.1	0.9978	8.6	0.99982
1.2	0.699	3.7	0.975	6.2	0.9980	8.7	0.99983
1.3	0.727	3.8	0.978	6.3	0.9982	8.8	0.99985
1.4	0.753	3.9	0.980	6.4	0.9983	8.9	0.99986
1.5	0.777	4.0	0.982	6.5	0.9985	9.0	0.99989
1.6	0.798	4.1	0.983	6.6	0.9986	9.1	0.99989
1.7	0.817	4.2	0.985	6.7	0.9988	9.2	0.99990
1.8	0.835	4.3	0.986	6.8	0.9989	9.3	0.99991
1.9	0.850	4.4	0.988	6.9	0.9990	9.4	0.99992
2.0	0.865	4.5	0.989	7.0	0.9991	9.5	0.99992
2.1	0.878	4.6	0.990	7.1	0.9992	9.6	0.99993
2.2	0.889	4.7	0.991	7.2	0.9993	9.7	0.99994
2.3	0.900	4.8	0.992	7.3	0.9993	9.8	0.99994
2.4	0.909	4.9	0.993	[illegible]	[illegible]		

ملحق (4) – توزيع ستيودنت t

العمود الأول خاص بدرجات الحرية (ν). عناوين بقية الأعمدة تعطي الاحتمالات (P) لقيمة t التي تزيد عن قيمة المدخل. استخدام التماثل لقيم t السالبة.



$p(t > t_0)$

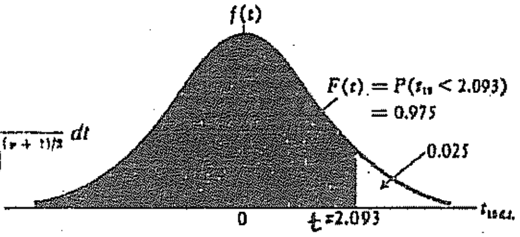
ν \ P	.10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

ملحق (5) – توزيع ستيودنت

t-distribution

$$F(t)=\int_{-\infty}^{t}\frac{\left(\frac{\nu-1}{2}\right)!}{\left(\frac{\nu-2}{2}\right)!\sqrt{\pi n}\left(1+\frac{t^2}{\nu}\right)^{(\nu+1)/2}}\,dt$$

Example: $n = 20$, $\nu = 19$



ν \ $F(t)$	$\alpha = .25$ .75	$\alpha = .10$ .90	$\alpha = .05$ .95	$\alpha = .025$ .975	$\alpha = .01$ .99	$\alpha = .005$ .995	$\alpha = .0005$ .9995
1	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.886	2.920	4.303	6.965	9.925	31.598
3	.765	1.638	2.353	3.182	4.541	5.841	12.941
4	.741	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	1.476	2.015	2.571	3.365	4.032	6.859
6	.718	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	1.415	1.895	2.365	2.998	3.499	5.405
8	.706	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	1.330	1.734	2.101	2.552	.2878	3.922
19	.688	1.328	1.729	2.093	2.539	2.861	3.883
20	.687	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	1.319	1.714	2.069	2.500	2.807	3.767
24	.685	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	1.314	1.703	2.052	2.473	2.771	3.690
28	.683	1.313	1.701	2.048	2.467	2.763	3.674
29	.683	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	3.551
60	.679	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	1.282	1.645	1.960	2.326	2.576	3.291

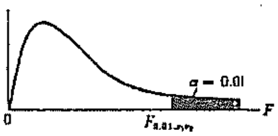
ملحق (6) – توزيع $F(\alpha = 0.01)$

F-distribution ($\alpha = 0.01$)

The following table gives the critical values of the *F* distribution for $\alpha = 0.01$. This probability represents the area exceeding the value of $F_{0.01,\nu_1,\nu_2}$, as shown by the shaded area in the figure below.

Examples: If $\nu_1 = 15$ (representing the greater mean square), and $\nu_2 = 20$, then the critical value for $\alpha = 0.01$ is 3.09.

$P(F > 3.09) = 0.01$,
$P(F < 3.09) = 0.99$.



VALUES OF $F_{0.01,\nu_1,\nu_2}$

ν_1 = Degrees of freedom for numerator

ν_2 = Degrees of freedom for denominator

ν_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	5.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	3.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

ملحق (7) – توزيع F$(\alpha = 0.05)$

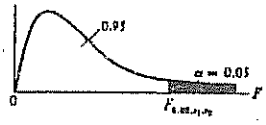
F-distribution ($\alpha = 0.05$)

Tables VI(a) and (b) from M. Merrington and C. M. Thompson, "Tables of percentage points of the inverted beta (F) distribution." *Biometrica*, Vol. 33 (1943) by permission of the *Biometrica* Trustees.

The following table gives the critical values of the F distribution for $\alpha = 0.05$. This probability represents the area exceeding the value of $F_{0.05,\nu_1,\nu_2}$, as shown by the shaded area in the figure below.

Examples: If $\nu_1 = 15$ (d.f. for the numerator), and $\nu_2 = 20$, then the critical value cutting off 0.05 is 2.20.

$P(F > 2.20) = 0.05,$
$P(F < 2.20) = 0.95.$



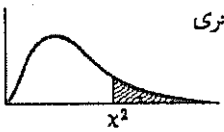
VALUES OF $F_{0.05,\nu_1,\nu_2}$

ν_1 = Degrees of freedom for numerator

ν_2 = Degrees of freedom for denominator

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	[illegible]	[illegible]

ملحق (8) – توزيع (كا^٢) χ^2



العمود الأول خاص بعدد درجات الحرية (ν). عناوين الأعمدة الأخرى تعطي الاحتمال (P) لقيمة χ^2 التي تزيد عن قيمة المدخل. لقيمة $100 < \nu$ عامل $\sqrt{2\chi^2} - \sqrt{2\nu - 1}$ كما لو كانت متغيراً معتدلاً قياسياً

ν \ P	0.995	0.975	0.050	0.025	0.010	0.005
1	$0.0^4 3927$	$0.0^3 9821$	3.84146	5.02389	6.63490	7.87944
2	0.010025	0.050636	5.99147	7.37776	9.21034	10.5966
3	0.071721	0.215795	7.81473	9.34840	11.3449	12.8381
4	0.206990	0.484419	9.48773	11.1433	13.2767	14.8602
5	0.411740	0.831211	11.0705	12.8325	15.0863	16.7496
6	0.675727	1.237347	12.5916	14.4494	16.8119	18.5476
7	0.989265	1.68987	14.0671	16.0128	18.4753	20.2777
8	1.344419	2.17973	15.5073	17.5346	20.0902	21.9550
9	1.734926	2.70039	16.9190	19.0228	21.6660	23.5893
10	2.15585	3.24697	18.3070	20.4831	23.2093	25.1882
11	2.60321	3.81575	19.6751	21.9200	24.7250	26.7569
12	3.07382	4.40379	21.0261	23.3367	26.2170	28.2995
13	3.56503	5.00874	22.3621	24.7356	27.6883	29.8194
14	4.07468	5.62872	23.6848	26.1190	29.1413	31.3193
15	4.60094	6.26214	24.9958	27.4884	30.5779	32.8013
16	5.14224	6.90766	26.2962	28.8454	31.9999	34.2672
17	5.69724	7.56418	27.5871	30.1910	33.4087	35.7185
18	6.26481	8.23075	28.8693	31.5264	34.8053	37.1564
19	6.84398	8.90655	30.1435	32.8523	36.1908	38.5822
20	7.43386	9.59083	31.4104	34.1696	37.5662	39.9968
21	8.03366	10.28293	32.6705	35.4789	38.9321	41.4010
22	8.64272	10.9823	33.9244	36.7807	40.2894	42.7956
23	9.26042	11.6885	35.1725	38.0757	41.6384	44.1813
24	9.88623	12.4001	36.4151	39.3641	42.9798	45.[illegible]
25	10.5197	13.1197	37.6525	40.6465	44.3141	46.9278
26	11.1603	13.8439	38.8852	41.9232	45.6417	48.2899
27	11.8076	14.5733	40.1133	43.1944	46.9630	49.6449
28	12.4613	15.3079	41.3372	44.4607	48.2782	50.9933
29	13.1211	16.0471	42.5569	45.7222	49.5879	52.3356
30	13.7867	16.7908	43.7729	46.9792	50.8922	53.6720
40	20.7065	24.4331	55.7585	59.3417	63.6907	66.7659
50	27.9907	32.3574	67.5048	71.4202	76.1539	79.4900
60	35.5346	40.4817	79.0819	83.2976	88.3794	91.9517
70	43.2752	48.7576	90.5312	95.0231	100.425	104.215
80	51.1720	57.1532	101.879	106.629	112.329	116.321
90	59.1963	65.6466	113.145	118.136	124.116	128.299
100	67.3276	74.2219	124.342	129.561	135.807	140.169

Poisson distribution (λ = 0.1 to λ = 4.0)

From *Handbook of Probability and Statistics* by R. S. Burington and D. C. May, Jr. Copyright 1953 by McGraw-Hill, Inc. Used with permission of McGraw-Hill Book Company.

The following table gives the probability of exactly x successes, for various values of λ, as defined by the Poisson mass function.

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Examples: If $\lambda = 1.5$, then $P(2) = 0.2510$, $P(3) = 0.1255$.

POISSON PROBABILITIES

x	λ = 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	.1438	.1647	.1839
3	.0002	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0002	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

x	λ = 1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.3662	.3614	.3543	.3452	.3347	.3230	.3106	.2975	.2842	.2707
2	.2014	.2169	.2303	.2417	.2510	.2584	.2640	.2678	.2700	.2707
3	.0738	.0867	.0998	.1128	.1255	.1378	.1496	.1607	.1710	.1804
4	.0203	.0260	.0324	.0395	.0471	.0551	.0636	.0723	.0812	.0902
5	.0045	.0062	.0084	.0111	.0141	.0176	.0216	.0260	.0309	.0361
6	.0008	.0012	.0018	.0026	.0035	.0047	.0061	.0078	.0098	.0120
7	.0001	.0002	.0003	.0005	.0008	.0011	.0015	.0020	.0027	.0034
8	.0000	.0000	.0001	.0001	.0001	.0002	.0003	.0005	.0006	.0009
9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002

x	λ = 2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.1225	.1108	.1003	.0907	.0821	.0743	.0672	.0608	.0550	.0498
1	.2572	.2438	.2306	.2177	.2052	.1931	.1815	.1703	.1596	.1494
2	.2700	.2681	.2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	.1890	.1966	.2033	.2090	.2138	.2176	.2205	.2225	.2237	.2240
4	.0992	.1082	.1169	.1254	.1336	.1414	.1488	.1557	.1622	.1680
5	.0417	.0476	.0538	.0602	.0668	.0735	.0804	.0872	.0940	.1008
6	.0146	.0174	.0206	.0241	.0278	.0319	.0362	.0407	.0455	.0504
7	.0044	.0055	.0068	.0083	.0099	.0118	.0139	.0163	.0188	.0216
8	.0011	.0015	.0019	.0025	.0031	.0038	.0047	.0057	.0068	.0081
9	.0003	.0004	.0005	.0007	.0009	.0011	.0014	.0018	.0022	.0027
10	.0001	.0001	.0001	.0002	.0002	.0003	.0004	.0005	.0006	.0008
11	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002	.0002
12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

x	λ = 3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1397	.1304	.1217	.1135	.1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	.2008	.1929	.1850	.1771	.1692	.1615	.1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	.1734	.1781	.1823	.1858	.1888	.1912	.1931	.1944	.1951	.1954
5	.1075	.1140	.1203	.1264	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	.0881	.0936	.0989	.1042
7	.0246	.0278	.0312	.0348	.0385	.0425	.0466	.0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	.0215	.0241	.0269	.0298
9	.0033	.0040	.0047	.0056	.0066	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0028	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0002	.0002
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

(Continued)

تابع ملحق (9)

Poisson distribution (λ = 8.1 to λ = 20)

x	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9.0
0	.0003	.0003	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
1	.0025	.0023	.0021	.0019	.0017	.0016	.0014	.0013	.0012	.0011
2	.0100	.0092	.0086	.0079	.0074	.0068	.0063	.0058	.0054	.0050
3	.0269	.0252	.0237	.0222	.0208	.0195	.0183	.0171	.0160	.0150
4	.0544	.0517	.0491	.0466	.0443	.0420	.0398	.0377	.0357	.0337
5	.0882	.0849	.0816	.0784	.0752	.0722	.0692	.0663	.0635	.0607
6	.1191	.1160	.1128	.1097	.1066	.1034	.1003	.0972	.0941	.0911
7	.1378	.1358	.1338	.1317	.1294	.1271	.1247	.1222	.1197	.1171
8	.1395	.1392	.1388	.1382	.1375	.1366	.1356	.1344	.1332	.1318
9	.1256	.1269	.1280	.1290	.1299	.1306	.1311	.1315	.1317	.1318
10	.1017	.1040	.1063	.1084	.1104	.1123	.1140	.1157	.1172	.1186
11	.0749	.0776	.0802	.0828	.0853	.0878	.0902	.0925	.0948	.0970
12	.0505	.0530	.0555	.0579	.0604	.0629	.0654	.0679	.0703	.0728
13	.0315	.0334	.0354	.0374	.0395	.0416	.0438	.0459	.0481	.0504
14	.0182	.0196	.0210	.0225	.0240	.0256	.0272	.0289	.0306	.0324
15	.0098	.0107	.0116	.0126	.0136	.0147	.0158	.0169	.0182	.0194
16	.0050	.0055	.0060	.0066	.0072	.0079	.0086	.0093	.0101	.0109
17	.0024	.0026	.0029	.0033	.0036	.0040	.0044	.0048	.0053	.0058
18	.0011	.0012	.0014	.0015	.0017	.0019	.0021	.0024	.0026	.0029
19	.0005	.0005	.0006	.0007	.0008	.0009	.0010	.0011	.0012	.0014
20	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0005	.0006
21	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0002	.0003
22	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001

x	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000
1	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005	.0005
2	.0046	.0043	.0040	.0037	.0034	.0031	.0029	.0027	.0025	.0023
3	.0140	.0131	.0123	.0115	.0107	.0100	.0093	.0087	.0081	.0076
4	.0319	.0302	.0285	.0269	.0254	.0240	.0226	.0213	.0201	.0189
5	.0581	.0555	.0530	.0506	.0483	.0460	.0439	.0418	.0398	.0378
6	.0881	.0851	.0822	.0793	.0764	.0736	.0709	.0682	.0656	.0631
7	.1145	.1118	.1091	.1064	.1037	.1010	.0982	.0955	.0928	.0901
8	.1302	.1286	.1269	.1251	.1232	.1212	.1191	.1170	.1148	.1126
9	.1317	.1315	.1311	.1306	.1300	.1293	.1284	.1274	.1263	.1251
10	.1198	.1210	.1219	.1228	.1235	.1241	.1245	.1249	.1250	.1251
11	.0991	.1012	.1031	.1049	.1067	.1083	.1098	.1112	.1125	.1137
12	.0752	.0776	.0799	.0822	.0844	.0866	.0888	.0908	.0928	.0948
13	.0526	.0549	.0572	.0594	.0617	.0640	.0662	.0685	.0707	.0729
14	.0342	.0361	.0380	.0399	.0419	.0439	.0459	.0479	.0500	.0521
15	.0208	.0221	.0235	.0250	.0265	.0281	.0297	.0313	.0330	.0347
16	.0118	.0127	.0137	.0147	.0157	.0168	.0180	.0192	.0204	.0217
17	.0063	.0069	.0075	.0081	.0088	.0095	.0103	.0111	.0119	.0128
18	.0032	.0035	.0039	.0042	.0046	.0051	.0055	.0060	.0065	.0071
19	.0015	.0017	.0019	.0021	.0023	.0026	.0028	.0031	.0034	.0037
20	.0007	.0008	.0009	.0010	.0011	.0012	.0014	.0015	.0017	.0019
21	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
22	.0001	.0001	.0002	.0002	.0002	.0002	.0003	.0003	.0004	.0004
23	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

x	11	12	13	14	15	16	17	18	19	20
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0010	.0004	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000
3	.0037	.0018	.0008	.0004	.0002	.0001	.0000	.0000	.0000	.0000
4	.0102	.0053	.0027	.0013	.0006	.0003	.0001	.0001	.0000	.0000
5	.0224	.0127	.0070	.0037	.0019	.0010	.0005	.0002	.0001	.0001
6	.0411	.0255	.0152	.0087	.0048	.0026	.0014	.0007	.0004	.0002
7	.0646	.0437	.0281	.0174	.0104	.0060	.0034	.0018	.0010	.0005
8	.0888	.0655	.0457	.0304	.0194	.0120	.0072	.0042	.0024	.0013
9	.1085	.0874	.0661	.0473	.0324	.0213	.0135	.0083	.0050	.0029
10	.1194	.1048	.0859	.0663	.0486	.0341	.0230	.0150	.0095	.0058
11	.1194	.1144	.1015	.0844	.0663	.0496	.0355	.0245	.0164	.0106
12	.1094	.1144	.1099	.0984	.0829	.0661	.0504	.0368	.0259	.0176
13	.0926	.1056	.1099	.1060	.0956	.0814	.0658	.0509	.0378	.0271
14	.0728	.0905	.1021	.1060	.1024	.0930	.0800	.0655	.0514	.0387
15	.0534	.0724	.0885	.0989	.1024	.0992	.0906	.0785	.0650	.0516
16	.0367	.0543	.0719	.0866	.0960	.0992	.0963	.0884	.0772	.0646
17	.0237	.0383	.0550	.0713	.0847	.0934	.0963	.0936	.0863	.0760
18	.0145	.0256	.0397	.0554	.0706	.0830	.0909	.0936	.0911	.0844
19	.0084	.0161	.0272	.0409	.0557	.0699	.0814	.0887	.0911	.0888
20	.0046	.0097	.0177	.0286	.0418	.0559	.0692	.0798	.0866	.0888
21	.0024	.0055	.0109	.0191	.0299	.0426	.0560	.0684	.0783	.0846
22	.0012	.0030	.0065	.0121	.0204	.0310	.0433	.0560	.0676	.0769
23	.0006	.0016	.0037	.0074	.0133	.0216	.0320	.0438	.0559	.0669
24	.0003	.0008	.0020	.0043	.0083	.0144	.0226	.0328	.0442	.0557
25	.0001	.0004	.0010	.0024	.0050	.0092	.0154	.0237	.0336	.0446
26	.0000	.0002	.0005	.0013	.0029	.0057	.0101	.0164	.0246	.0343
27	.0000	.0001	.0002	.0007	.0016	.0034	.0063	.0109	.0173	.0254
28	.0000	.0000	.0001	.0003	.0009	.0019	.0038	.0070	.0117	.0181
29	.0000	.0000	.0001	.0002	.0004	.0011	.0023	.0044	.0077	.0125
30	.0000	.0000	.0000	.0001	.0002	.0006	.0013	.0026	.0049	.0083
31	.0000	.0000	.0000	.0000	.0001	.0003	.0007	.0015	.0030	.0054
32	.0000	.0000	.0000	.0000	.0001	.0001	.0004	.0009	.0018	.0034
33	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0005	.0010	.0020
34	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0012
35	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0007
36	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0004
37	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
38	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
39	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

تابع ملحق (9)

Poisson distribution (λ = 4.1 to λ = 8.0)

x	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	.0166	.0150	.0136	.0123	.0111	.0101	.0091	.0082	.0074	.0067
1	.0679	.0630	.0583	.0540	.0500	.0462	.0427	.0395	.0365	.0337
2	.1393	.1323	.1254	.1188	.1125	.1063	.1005	.0948	.0894	.0842
3	.1904	.1852	.1798	.1743	.1687	.1631	.1574	.1517	.1460	.1404
4	.1951	.1944	.1933	.1917	.1898	.1875	.1849	.1820	.1789	.1755
5	.1600	.1633	.1662	.1687	.1708	.1725	.1738	.1747	.1753	.1755
6	.1093	.1143	.1191	.1237	.1281	.1323	.1362	.1398	.1432	.1462
7	.0640	.0686	.0732	.0778	.0824	.0869	.0914	.0959	.1002	.1044
8	.0328	.0360	.0393	.0428	.0463	0500	.0537	.0575	.0614	.0653
9	.0150	.0168	.0188	.0209	.0232	.0255	.0280	.0307	.0334	.0363
10	.0061	.0071	.0081	.0092	.0104	.0118	.0132	.0147	.0164	.0181
11	.0023	.0027	.0032	.0037	.0043	.0049	.0056	.0064	.0073	.0082
12	.0008	.0009	.0011	.0014	.0016	.0019	.0022	.0026	.0030	.0034
13	.0002	.0003	.0004	.0005	.0006	.0007	.0008	.0009	.0011	.0013
14	.0001	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005
15	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0002

x	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	.0061	.0055	.0050	.0045	.0041	.0037	.0033	.0030	.0027	.0025
1	.0311	.0287	.0265	.0244	.0225	.0207	.0191	.0176	.0162	.0149
2	.0793	.0746	.0701	.0659	.0618	.0580	.0544	.0509	.0477	.0446
3	.1348	.1293	.1239	.1185	.1133	.1082	.1033	.0985	.0938	.0892
4	.1719	.1681	.1641	.1600	.1558	.1515	.1472	.1428	.1383	.1339
5	.1753	.1748	.1740	.1728	.1714	.1697	.1678	.1620	.1632	.1606
6	.1490	.1515	.1537	.1555	.1571	.1584	.1594	.1656	.1605	.1606
7	.1086	.1125	.1163	.1200	.1234	.1267	.1298	.1301	.1353	.1377
8	.0692	.0731	.0771	.0810	.0849	.0887	.0925	.0925	.0998	.1033
9	.0392	.0423	0.454	.0486	.0519	.0552	.0586	.0662	.0654	.0688
10	.0200	.0320	.0241	.0262	.0285	.0309	.0334	.0359	.0386	.0413
11	.0093	.0104	.0116	.0129	.0143	.0157	.0173	.0190	.0207	.0225
12	.0039	.0045	.0051	.0058	.0065	.0073	.0082	.0092	.0102	.0113
13	.0015	.0018	.0021	.0024	.0028	.0032	.0036	.0041	.0046	.0052
14	.0006	.0007	.0008	.0009	.0011	.0013	.0015	.0017	.0019	.0022
15	.0002	.0002	.0003	.0003	.0004	.0005	.0006	.0007	.0008	.0009
16	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003
17	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001

x	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	.0022	.0020	.0018	.0017	.0015	.0014	.0012	.0011	.0010	.0009
1	.0137	.0126	.0116	.0106	.0098	.0090	.0082	.0076	.0070	.0064
2	.0417	.0390	.0364	.0340	.0318	.0296	.0276	.0258	.0240	.0223
3	.0848	.0806	.0765	.0726	.0688	.0652	.0617	.0584	.0552	.0521
4	.1294	.1249	.1205	.1162	.1118	.1076	.1034	.0992	.0952	.0912
5	.1579	.1549	.1519	.1487	.1454	.1420	.1385	.1349	.1314	.1277
6	.1605	.1601	.1595	.1586	.1575	.1562	.1546	.1529	.1511	.1490
7	.1399	.1418	.1435	.1450	.1462	.1472	.1480	.1486	.1489	.1490
8	.1066	.1099	.1130	.1160	.1188	.1215	.1240	.1263	.1284	.1304
9	.0723	.0757	.0791	.0825	.0858	.0891	.0923	.0954	.0985	.1014
10	.0441	.0469	.0498	.0528	.0558	.0588	.0618	.0649	.0679	.0710
11	.0245	.0265	.0285	.0307	.0330	.0353	.0377	.0401	.0426	.0452
12	.0124	.0137	.0150	.0164	.0179	.0194	.0210	.0227	.0245	.0264
13	.0058	0.065	.0073	.0081	.0089	.0098	.0108	.0119	.0130	.0142
14	.0025	.0029	.0033	.0037	.0041	.0046	.0052	.0058	.0064	.0071
15	.0010	.0012	.0014	.0016	.0018	.0020	.0023	.0026	.0029	.0033
16	.0004	.0005	.0005	.0006	.0007	.0008	.0010	.0011	.0013	.0014
17	.0001	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0006
18	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002
19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

x	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	.0008	.0007	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0003
1	.0059	.0054	.0049	.0045	.0041	.0038	.0035	.0032	.0029	.0027
2	.0208	.0194	.0180	.0167	.0156	.0145	.0134	.0125	.0116	.0107
3	.0492	.0464	.0438	.0413	.0389	.0366	.0345	.0324	.0305	.0286
4	.0874	.0836	.0799	.0764	.0729	.0696	.0663	.0632	.0602	.0573
5	.1241	.1204	.1167	.1130	.1094	.1057	.1021	.0986	.0951	.0916
6	.1468	.1445	.1420	.1394	.1367	.1339	.1311	.1282	.1252	.1221
7	.1489	.1486	.1481	.1474	.1465	.1454	.1442	.1428	.1413	.1396
8	.1321	.1337	.1351	.1363	.1373	.1382	.1388	.1392	.1395	.1396
9	.1042	.1070	.1096	.1121	.1144	.1167	.1187	.1207	.1224	.1241
10	.0740	.0770	.0800	.0829	.0858	.0887	.0914	.0941	.0967	.0993
11	.0478	.0504	.0531	.0558	.0585	.0613	.0640	.0667	.0695	.0722
12	.0283	.0303	.0323	.0344	.0366	.0388	.0411	.0434	.0457	.0481
13	.0154	.0168	.0181	.0196	.0211	.0227	.0243	.0260	.0278	.0296
14	.0078	.0086	.0095	.0104	.0113	.0123	.0134	.0145	.0157	.0169
15	.0037	.0041	.0046	.0051	.0057	.0062	.0069	.0075	.0083	.0090
16	.0016	.0019	.0021	.0024	.0026	.0030	.0033	.0037	.0041	.0045
17	.0007	.0008	.0009	.0010	.0012	.0013	.0015	.0017	.1119	.0021
18	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
19	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003	.0003	.0004
20	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002
21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001

(Continued)

ملحق (10) التوزيع الثنائي

Binomial distribution ($n = 1, n = 2$)

From Robert Schlaifer, *Analysis of Decisions Under Uncertainty* (New York: McGraw-Hill Book Co., Inc., Preliminary Edition, Volume II, 1967) by specific permission of the President and Fellows of Harvard College, who hold the copyright.

The following table gives values of the binomial mass function defined by

$$P(x) = {}_nC_x\, p^x(1-p)^{n-x}$$

$$= \frac{n!}{x!(n-x)!}\, p^x(1-p)^{n-x}.$$

This is the probability of exactly x successes in n independent Bernoulli trials with probability of success on a single trial equal to p. The values of x at the left of any section are to be used in conjunction with the values of p at the top of that section; the values of x at the right of any section are to be used in conjunction with the values of p at the bottom of that section.

Example: To evaluate $P(x)$ for $n = 5$, $x = 3$, and $p = 0.83$, locate the section of the table for $n = 5$, the column for $p = 0.83$, and the row for $x = 3$, and read

$$P(x) = 0.1632.$$

$n = 1$

x	p	01	02	03	04	05	06	07	08	09	10		
0		9900	9800	9700	9600	9500	9400	9300	9200	9100	9000		1
1		0100	0200	0300	0400	0500	0600	0700	0800	0900	1000		0
		99	98	97	96	95	94	93	92	91	90	p	x

x	p	11	12	13	14	15	16	17	18	19	20		
0		8900	8800	8700	8600	8500	8400	8300	8200	8100	8000		1
1		1100	1200	1300	1400	1500	1600	1700	1800	1900	2000		0
		89	88	87	86	85	84	83	82	81	80	p	x

x	p	21	22	23	24	25	26	27	28	29	30		
0		7900	7800	7700	7600	7500	7400	7300	7200	7100	7000		1
1		2100	2200	2300	2400	2500	2600	2700	2800	2900	3000		0
		79	78	77	76	75	74	73	72	71	70	p	x

x	p	31	32	33	34	35	36	37	38	39	40		
0		6900	6800	6700	6600	6500	6400	6300	6200	6100	6000		1
1		3100	3200	3300	3400	3500	3600	3700	3800	3900	4000		0
		69	68	67	66	65	64	63	62	61	60	p	x

x	p	41	42	43	44	45	46	47	48	49	50		
0		5900	5800	5700	5600	5500	5400	5300	5200	5100	5000		1
1		4100	4200	4300	4400	4500	4600	4700	4800	4900	5000		0
		59	58	57	56	55	54	53	52	51	50	p	x

$n = 2$

x	p	01	02	03	04	05	06	07	08	09	10	
0		9801	9604	9409	9216	9025	8836	8649	8464	8281	8100	
1		0198	0392	0582	0768	0950	1128	1302	1472	1638	1800	
2		0001	0004	0009	0016	0025	0036	0049	0064	0081	0100	
		99	98	97	96	95	94	93	92	91	90	p

x	p	11	12	13	14	15	16	17	18	19	20	
0		7921	7744	7569	7396	7225	7056	6889	6724	6561	6400	
1		1958	2112	2262	2408	2550	2688	2822	2952	3078	3200	
2		0121	0144	0169	0196	0225	0256	0289	0324	0361	0400	
		89	88	87	86	85	84	83	82	81	80	p

x	p	21	22	23	24	25	26	27	28	29	30	
0		6241	6084	5929	5776	5625	5476	5329	5184	5041	4900	
1		3318	3432	3542	3648	3750	3848	3942	4032	4118	4200	
2		0441	0484	0529	0576	0625	0676	0729	0784	0841	0900	
		79	78	77	76	75	74	73	72	71	70	p

x	p	31	32	33	34	35	36	37	38	39	40	
0		4761	4624	4489	4356	4225	4096	3969	3844	3721	3600	
1		4278	4352	4422	4488	4550	4608	4662	4712	4758	4800	
2		0961	1024	1089	1156	1225	1296	1369	1444	1521	1600	
		69	68	67	66	65	64	63	62	61	60	p

x	p	41	42	43	44	45	46	47	48	49	50	
0		3481	3364	3249	3136	3025	2916	2809	2704	2601	2500	
1		4838	4872	4902	4928	4950	4968	4982	4992	4998	5000	
2		1681	1764	1849	1936	2025	2116	2209	2304	2401	2500	
		59	58	57	56	55	54	53	52	51	50	p

(Continu

تابع ملحق (10)

Binomial distribution ($n = 3$, $n = 4$)

$n = 3$

x \ p	01	02	03	04	05	06	07	08	09	10	
0	9703	9412	9127	8847	8574	8306	8044	7787	7536	7290	3
1	0294	0576	0847	1106	1354	1590	1816	2031	2236	2430	2
2	0003	0012	0026	0046	0071	0102	0137	0177	0221	0270	1
3	0000	0000	0000	0001	0001	0002	0003	0005	0007	0010	0
	99	98	97	96	95	94	93	92	91	90	p \ x

x \ p	11	12	13	14	15	16	17	18	19	20	
0	7050	6815	6585	6361	6141	5927	5718	5514	5314	5120	3
1	2614	2788	2952	3106	3251	3387	3513	3631	3740	3840	2
2	0323	0380	0441	0506	0574	0645	0720	0797	0877	0960	1
3	0013	0017	0022	0027	0034	0041	0049	0058	0069	0080	0
	89	88	87	86	85	84	83	82	81	80	p \ x

x \ p	21	22	23	24	25	26	27	28	29	30	
0	4930	4746	4565	4390	4219	4052	3890	3732	3579	3430	3
1	3932	4014	4091	4159	4219	4271	4316	4355	4386	4410	2
2	1045	1133	1222	1313	1406	1501	1597	1693	1791	1890	1
3	0093	0106	0122	0138	0156	0176	0197	0220	0244	0270	0
	79	78	77	76	75	74	73	72	71	70	p \ x

x \ p	31	32	33	34	35	36	37	38	39	40	
0	3285	3144	3008	2875	2746	2621	2500	2383	2270	2160	3
1	4428	4439	4444	4443	4436	4424	4406	4382	4354	4320	2
2	1989	2089	2189	2289	2389	2488	2587	2686	2783	2880	1
3	0298	0328	0359	0393	0429	0467	0507	0549	0593	0640	0
	69	68	67	66	65	64	63	62	61	60	p \ x

x \ p	41	42	43	44	45	46	47	48	49	50	
0	2054	1951	1852	1756	1664	1575	1489	1406	1327	1250	3
1	4282	4239	4191	4140	4084	4024	3961	3894	3823	3750	2
2	2975	3069	3162	3252	3341	3428	3512	3594	3674	3750	1
3	0689	0741	0795	0852	0911	0973	1038	1106	1176	1250	0
	59	58	57	56	55	54	53	52	51	50	p \ x

$n = 4$

x \ p	01	02	03	04	05	06	07	08	09	10	
0	9606	9224	8853	8493	8145	7807	7481	7164	6857	6561	4
1	0388	0753	1095	1416	1715	1993	2252	2492	2713	2916	3
2	0006	0023	0051	0088	0135	0191	0254	0325	0402	0486	2
3	0000	0000	0001	0002	0005	0008	0013	0019	0027	0036	1
4	0000	0000	0000	0000	0000	0000	0000	0000	0001	0001	0
	99	98	97	96	95	94	93	92	91	90	p \ x

x \ p	11	12	13	14	15	16	17	18	19	20	
0	6274	5997	5729	5470	5220	4979	4746	4521	4305	4096	4
1	3102	3271	3424	3562	3685	3793	3888	3970	4039	4096	3
2	0575	0669	0767	0870	0975	1084	1195	1307	1421	1536	2
3	0047	0061	0076	0094	0115	0138	0163	0191	0222	0256	1
4	0001	0002	0003	0004	0005	0007	0008	0010	0013	0016	0
	89	88	87	86	85	84	83	82	81	80	p \ x

x \ p	21	22	23	24	25	26	27	28	29	30	
0	3895	3702	3515	3336	3164	2999	2840	2687	2541	2401	4
1	4142	4176	4200	4214	4219	4214	4201	4180	4152	4116	3
2	1651	1767	1882	1996	2109	2221	2331	2439	2544	2646	2
3	0293	0332	0375	0420	0469	0520	0575	0632	0693	0756	1
4	0019	0023	0028	0033	0039	0046	0053	0061	0071	0081	0
	79	78	77	76	75	74	73	72	71	70	p \ x

x \ p	31	32	33	34	35	36	37	38	39	40	
0	2267	2138	2015	1897	1785	1678	1575	1478	1385	1296	4
1	4074	4025	3970	3910	3845	3775	3701	3623	3541	3456	3
2	2745	2841	2933	3021	3105	3185	3260	3330	3396	3456	2
3	0822	0891	0963	1038	1115	1194	1276	1361	1447	1536	1
4	0092	0105	0119	0134	0150	0168	0187	0209	0231	0256	0
	69	68	67	66	65	64	63	62	61	60	p \ x

x \ p	41	42	43	44	45	46	47	48	49	50	
0	1212	1132	1056	0983	0915	0850	0789	0731	0677	0625	4
1	3368	3278	3185	3091	2995	2897	2799	2700	2600	2500	3
2	3511	3560	3604	3643	3675	3702	3723	3738	3747	3750	2
3	1627	1719	1813	1908	2005	2102	2201	2300	2400	2500	1
4	0283	0311	0342	0375	0410	0448	0488	0531	0576	0625	0
	59	58	57	56	55	54	53	52	51	50	p \ x

(Continued)

تابع ملحق (10)

Binomial distribution ($n = 5$, $n = 6$)

$n = 5$

x	p	01	02	03	04	05	06	07	08	09	10		
0		9510	9039	8587	8154	7738	7339	6957	6591	6240	5905		5
1		0480	0922	1328	1699	2036	2342	2618	2866	3086	3280		4
2		0010	0038	0082	0142	0214	0299	0394	0498	0610	0729		3
3		0000	0001	0003	0006	0011	0019	0030	0043	0060	0081		2
4		0000	0000	0000	0000	0000	0001	0001	0002	0003	0004		1
		99	98	97	96	95	94	93	92	91	90	p	x

x	p	11	12	13	14	15	16	17	18	19	20		
0		5584	5277	4984	4704	4437	4182	3939	3707	3487	3277		5
1		3451	3598	3724	3829	3915	3983	4034	4069	4089	4096		4
2		0853	0981	1113	1247	1382	1517	1652	1786	1919	2048		3
3		0105	0134	0166	0203	0244	0289	0338	0392	0450	0512		2
4		0007	0009	0012	0017	0022	0028	0035	0043	0053	0064		1
5		0000	0000	0000	0001	0001	0001	0001	0002	0002	0003		0
		89	88	87	86	85	84	83	82	81	80	p	x

x	p	21	22	23	24	25	26	27	28	29	30		
0		3077	2887	2707	2536	2373	2219	2073	1935	1804	1681		5
1		4090	4072	4043	4003	3955	3898	3834	3762	3685	3601		4
2		2174	2297	2415	2529	2637	2739	2836	2926	3010	3087		3
3		0578	0648	0721	0798	0879	0962	1049	1138	1229	1323		2
4		0077	0091	0108	0125	0146	0169	0194	0221	0251	0283		1
5		0004	0005	0006	0008	0010	0012	0014	0017	0021	0024		0
		79	78	77	76	75	74	73	72	71	70	p	x

x	p	31	32	33	34	35	36	37	38	39	40		
0		1564	1454	1350	1252	1160	1074	0992	0916	0845	0778		5
1		3513	3421	3325	3226	3124	3020	2914	2808	2700	2592		4
2		3157	3220	3275	3323	3364	3397	3423	3441	3452	3456		3
3		1418	1515	1613	1712	1811	1911	2010	2109	2207	2304		2
4		0319	0357	0397	0441	0488	0537	0590	0646	0706	0768		1
5		0029	0034	0039	0045	0053	0060	0069	0079	0090	0102		0
		69	68	67	66	65	64	63	62	61	60	p	x

x	p	41	42	43	44	45	46	47	48	49	50		
0		0715	0656	0602	0551	0503	0459	0418	0380	0345	0312		5
1		2484	2376	2270	2164	2059	1956	1854	1755	1657	1562		4
2		3452	3442	3424	3400	3369	3332	3289	3240	3185	3125		3
3		2399	2492	2583	2671	2757	2838	2916	2990	3060	3125		2
4		0834	0902	0974	1049	1128	1209	1293	1380	1470	1562		1
5		0116	0131	0147	0165	0185	0206	0229	0255	0282	0312		0
		59	58	57	56	55	54	53	52	51	50	p	x

$n = 6$

x	p	01	02	03	04	05	06	07	08	09	10		
0		9415	8858	8330	7828	7351	6899	6470	6064	5679	5314		6
1		0571	1085	1546	1957	2321	2642	2922	3164	3370	3543		5
2		0014	0055	0120	0204	0305	0422	0550	0688	0833	0984		4
3		0000	0002	0005	0011	0021	0036	0055	0080	0110	0146		3
4		0000	0000	0000	0000	0001	0002	0003	0005	0008	0012		2
5		0000	0000	0000	0000	0000	0000	0000	0000	0000	0001		1
		99	98	97	96	95	94	93	92	91	90	p	x

x	p	11	12	13	14	15	16	17	18	19	20		
0		4970	4644	4336	4046	3771	3513	3269	3040	2824	2621		6
1		3685	3800	3888	3952	3993	4015	4018	4004	3975	3932		5
2		1139	1295	1452	1608	1762	1912	2057	2197	2331	2458		4
3		0188	0236	0289	0349	0415	0486	0563	0643	0729	0819		3
4		0017	0024	0032	0043	0055	0069	0086	0106	0128	0154		2
5		0001	0001	0002	0003	0004	0005	0007	0009	0012	0015		1
6		0000	0000	0000	0000	0000	0000	0000	0000	0000	0001		0
		89	88	87	86	85	84	83	82	81	80	p	x

x	p	21	22	23	24	25	26	27	28	29	30		
0		2431	2252	2084	1927	1780	1642	1513	1393	1281	1176		6
1		3877	3811	3735	3651	3560	3462	3358	3251	3139	3025		5
2		2577	2687	2789	2882	2966	3041	3105	3160	3206	3241		4
3		0913	1011	1111	1214	1318	1424	1531	1639	1746	1852		3
4		0182	0214	0249	0287	0330	0375	0425	0478	0535	0595		2
5		0019	0024	0030	0036	0044	0053	0063	0074	0087	0102		1
6		0001	0001	0001	0002	0002	0003	0004	0005	0006	0007		0
		79	78	77	76	75	74	73	72	71	70	p	x

x	p	31	32	33	34	35	36	37	38	39	40		
0		1079	0989	0905	0827	0754	0687	0625	0568	0515	0467		6
1		2909	2792	2673	2555	2437	2319	2203	2089	1976	1866		5
2		3267	3284	3292	3290	3280	3261	3235	3201	3159	3110		4
3		1957	2061	2162	2260	2355	2446	2533	2616	2693	2765		3
4		0660	0727	0799	0873	0951	1032	1116	1202	1291	1382		2
5		0119	0137	0157	0180	0205	0232	0262	0295	0330	0369		1
6		0009	0011	0013	0015	0018	0022	0026	0030	0035	0041		0
		69	68	67	66	65	64	63	62	61	60	p	x

x	p	41	42	43	44	45	46	47	48	49	50		
0		0422	0381	0343	0308	0277	0248	0222	0198	0176	0156		6
1		1759	1654	1552	1454	1359	1267	1179	1095	1014	0937		5
2		3055	2994	2928	2856	2780	2699	2615	2527	2436	2344		4
3		2831	2891	2945	2992	3032	3065	3091	3110	3121	3125		3
4		1475	1570	1666	1763	1861	1958	2056	2153	2249	2344		2
5		0410	0455	0503	0554	0609	0667	0729	0795	0864	0937		1
6		0048	0055	0063	0073	0083	0095	0108	0122	0138	0156		0
		59	58	57	56	55	54	53	52	51	50	p	x

(Continued)

تابع ملحق (10)

Binomial distribution ($n = 7$, $n = 8$)

$n = 7$

x	p	01	02	03	04	05	06	07	08	09	10		
0		9321	8681	8080	7514	6983	6485	6017	5578	5168	4783		7
1		0659	1240	1749	2192	2573	2897	3170	3396	3578	3720		6
2		0020	0076	0162	0274	0406	0555	0716	0886	1061	1240		5
3		0000	0003	0008	0019	0036	0059	0090	0128	0175	0230		4
4		0000	0000	0000	0001	0002	0004	0007	0011	0017	0026		3
5		0000	0000	0000	0000	0000	0000	0000	0001	0001	0002		2
		99	98	97	96	95	94	93	92	91	90	p	x

x	p	11	12	13	14	15	16	17	18	19	20		
0		4423	4087	3773	3479	3206	2951	2714	2493	2288	2097		7
1		3827	3901	3946	3965	3960	3935	3891	3830	3756	3670		6
2		1419	1596	1769	1936	2097	2248	2391	2523	2643	2753		5
3		0292	0363	0441	0525	0617	0714	0816	0923	1033	1147		4
4		0036	0049	0066	0086	0109	0136	0167	0203	0242	0287		3
5		0003	0004	0006	0008	0012	0016	0021	0027	0034	0043		2
6		0000	0000	0000	0000	0001	0001	0001	0002	0003	0004		1
		89	88	87	86	85	84	83	82	81	80	p	x

x	p	21	22	23	24	25	26	27	28	29	30		
0		1920	1757	1605	1465	1335	1215	1105	1003	0910	0824		7
1		3573	3468	3356	3237	3115	2989	2860	2731	2600	2471		6
2		2850	2935	3007	3067	3115	3150	3174	3186	3186	3177		5
3		1263	1379	1497	1614	1730	1845	1956	2065	2169	2269		4
4		0336	0389	0447	0510	0577	0648	0724	0803	0886	0972		3
5		0054	0066	0080	0097	0115	0137	0161	0187	0217	0250		2
6		0005	0006	0008	0010	0013	0016	0020	0024	0030	0036		1
7		0000	0000	0000	0000	0001	0001	0001	0001	0002	0002		0
		79	78	77	76	75	74	73	72	71	70	p	x

x	p	31	32	33	34	35	36	37	38	39	40		
0		0745	0672	0606	0546	0490	0440	0394	0352	0314	0280		7
1		2342	2215	2090	1967	1848	1732	1619	1511	1407	1306		6
2		3156	3127	3088	3040	2985	2922	2853	2778	2698	2613		5
3		2363	2452	2535	2610	2679	2740	2793	2838	2875	2903		4
4		1062	1154	1248	1345	1442	1541	1640	1739	1838	1935		3
5		0286	0326	0369	0416	0466	0520	0578	0640	0705	0774		2
6		0043	0051	0061	0071	0084	0098	0113	0131	0150	0172		1
7		0003	0003	0004	0005	0006	0008	0009	0011	0014	0016		0
		69	68	67	66	65	64	63	62	61	60	p	x

x	p	41	42	43	44	45	46	47	48	49	50		
0		0249	0221	0195	0173	0152	0134	0117	0103	0090	0078		7
1		1211	1119	1032	0950	0872	0798	0729	0664	0604	0547		6
2		2524	2431	2336	2239	2140	2040	1940	1840	1740	1641		5
3		2923	2934	2937	2932	2918	2897	2867	2830	2786	2734		4
4		2031	2125	2216	2304	2388	2468	2543	2612	2676	2734		3
5		0847	0923	1003	1086	1172	1261	1353	1447	1543	1641		2
6		0196	0223	0252	0284	0320	0358	0400	0445	0494	0547		1
7		0019	0023	0027	0032	0037	0044	0051	0059	0068	0078		0
		59	58	57	56	55	54	53	52	51	50	p	x

$n = 8$

x	p	01	02	03	04	05	06	07	08	09	10		
0		9227	8508	7837	7214	6634	6096	5596	5132	4703	4305		8
1		0746	1389	1939	2405	2793	3113	3370	3570	3721	3826		7
2		0026	0099	0210	0351	0515	0695	0888	1087	1288	1488		6
3		0001	0004	0013	0029	0054	0089	0134	0189	0255	0331		5
4		0000	0000	0001	0002	0004	0007	0013	0021	0031	0046		4
5		0000	0000	0000	0000	0000	0000	0001	0001	0002	0004		3
		99	98	97	96	95	94	93	92	91	90	p	x

x	p	11	12	13	14	15	16	17	18	19	20		
0		3937	3596	3282	2992	2725	2479	2252	2044	1853	1678		8
1		3892	3923	3923	3897	3847	3777	3691	3590	3477	3355		7
2		1684	1872	2052	2220	2376	2518	2646	2758	2855	2936		6
3		0416	0511	0613	0723	0839	0959	1084	1211	1339	1468		5
4		0064	0087	0115	0147	0185	0228	0277	0332	0393	0459		4
5		0006	0009	0014	0019	0026	0035	0045	0058	0074	0092		3
6		0000	0001	0001	0002	0002	0003	0005	0006	0009	0011		2
7		0000	0000	0000	0000	0000	0000	0000	0000	0001	0001		1
		89	88	87	86	85	84	83	82	81	80	p	x

x	p	21	22	23	24	25	26	27	28	29	30		
0		1517	1370	1236	1113	1001	0899	0806	0722	0646	0576		8
1		3226	3092	2953	2812	2670	2527	2386	2247	2110	1977		7
2		3002	3052	3087	3108	3115	3108	3089	3058	3017	2965		6
3		1596	1722	1844	1963	2076	2184	2285	2379	2464	2541		5
4		0530	0607	0689	0775	0865	0959	1056	1156	1258	1361		4
5		0113	0137	0165	0196	0231	0270	0313	0360	0411	0467		3
6		0015	0019	0025	0031	0038	0047	0058	0070	0084	0100		2
7		0001	0002	0002	0003	0004	0005	0006	0008	0010	0012		1
8		0000	0000	0000	0000	0000	0000	0000	0000	0001	0001		0
		79	78	77	76	75	74	73	72	71	70	p	x

x	p	31	32	33	34	35	36	37	38	39	40		
0		0514	0457	0406	0360	0319	0281	0248	0218	0192	0168		8
1		1847	1721	1600	1484	1373	1267	1166	1071	0981	0896		7
2		2904	2835	2758	2675	2587	2494	2397	2297	2194	2090		6
3		2609	2668	2717	2756	2786	2805	2815	2815	2806	2787		5
4		1465	1569	1673	1775	1875	1973	2067	2157	2242	2322		4
5		0527	0591	0659	0732	0808	0888	0971	1058	1147	1239		3
6		0118	0139	0162	0188	0217	0250	0285	0324	0367	0413		2
7		0015	0019	0023	0028	0033	0040	0048	0057	0067	0079		1
8		0001	0001	0001	0002	0002	0003	0004	0004	0005	0007		0
		69	68	67	66	65	64	63	62	61	60	p	x

x	p	41	42	43	44	45	46	47	48	49	50		
0		0147	0128	0111	0097	0084	0072	0062	0053	0046	0039		8
1		0816	0742	0672	0608	0548	0493	0442	0395	0352	0312		7
2		1985	1880	1776	1672	1569	1469	1371	1275	1183	1094		6
3		2759	2723	2679	2627	2568	2503	2431	2355	2273	2187		5
4		2397	2465	2526	2580	2627	2665	2695	2717	2730	2734		4
5		1332	1428	1525	1622	1719	1816	1912	2006	2098	2187		3
6		0463	0517	0575	0637	0703	0774	0848	0926	1008	1094		2
7		0092	0107	0124	0143	0164	0188	0215	0244	0277	0312		1
8		0008	0010	0012	0014	0017	0020	0024	0028	0033	0039		0
		59	58	57	56	55	54	53	52	51	50	p	x

(Continued)

تابع ملحق (10)

n = 9

x	p = 01	02	03	04	05	06	07	08	09	10	
0	9135	8337	7602	6925	6302	5730	5204	4722	4279	3874	9
1	0830	1531	2116	2597	2985	3292	3525	3695	3809	3874	8
2	0034	0125	0262	0433	0629	0840	1061	1285	1507	1722	7
3	0001	0006	0019	0042	0077	0125	0186	0261	0348	0446	6
4	0000	0000	0001	0003	0006	0012	0021	0034	0052	0074	5
5	0000	0000	0000	0000	0000	0001	0002	0003	0005	0008	4
6	0000	0000	0000	0000	0000	0000	0000	0000	0000	0001	3
	99	98	97	96	95	94	93	92	91	90 = p	x

x	p = 11	12	13	14	15	16	17	18	19	20	
0	3504	3165	2855	2573	2316	2082	1869	1676	1501	1342	9
1	3897	3884	3840	3770	3679	3569	3446	3312	3169	3020	8
2	1927	2119	2295	2455	2597	2720	2823	2908	2973	3020	7
3	0556	0674	0800	0933	1069	1209	1349	1489	1627	1762	6
4	0103	0138	0179	0228	0283	0345	0415	0490	0573	0661	5
5	0013	0019	0027	0037	0050	0066	0085	0108	0134	0165	4
6	0001	0002	0003	0004	0006	0008	0012	0016	0021	0028	3
7	0000	0000	0000	0000	0000	0001	0001	0001	0002	0003	2
	89	88	87	86	85	84	83	82	81	80 = p	x

x	p = 21	22	23	24	25	26	27	28	29	30	
0	1199	1069	0952	0846	0751	0665	0589	0520	0458	0404	9
1	2867	2713	2558	2404	2253	2104	1960	1820	1685	1556	8
2	3049	3061	3056	3037	3003	2957	2899	2831	2754	2668	7
3	1891	2014	2130	2238	2336	2424	2502	2569	2624	2668	6
4	0754	0852	0954	1060	1168	1278	1388	1499	1608	1715	5
5	0200	0240	0285	0335	0389	0447	0513	0583	0657	0735	4
6	0036	0045	0057	0070	0087	0105	0127	0151	0179	0210	3
7	0004	0005	0007	0010	0012	0016	0020	0025	0031	0039	2
8	0000	0000	0001	0001	0001	0001	0002	0002	0003	0004	1
	79	78	77	76	75	74	73	72	71	70 = p	x

x	p = 31	32	33	34	35	36	37	38	39	40	
0	0355	0311	0272	0238	0207	0180	0156	0135	0117	0101	9
1	1433	1317	1206	1102	1004	0912	0826	0747	0673	0605	8
2	2576	2478	2376	2270	2162	2052	1941	1831	1721	1612	7
3	2701	2721	2731	2729	2716	2693	2660	2618	2567	2508	6
4	1820	1921	2017	2109	2194	2272	2344	2407	2462	2508	5
5	0818	0904	0994	1086	1181	1278	1376	1475	1574	1672	4
6	0245	0284	0326	0373	0424	0479	0539	0603	0671	0743	3
7	0047	0057	0069	0082	0098	0116	0136	0158	0184	0212	2
8	0005	0007	0008	0011	0013	0016	0020	0024	0029	0035	1
9	0000	0000	0000	0001	0001	0001	0001	0002	0002	0003	0
	69	68	67	66	65	64	63	62	61	60 = p	x

x	p = 41	42	43	44	45	46	47	48	49	50	
0	0087	0074	0064	0054	0046	0039	0033	0028	0023	0020	9
1	0542	0484	0431	0383	0339	0299	0263	0231	0202	0176	8
2	1506	1402	1301	1204	1110	1020	0934	0853	0776	0703	7
3	2442	2369	2291	2207	2119	2027	1933	1837	1739	1641	6
4	2545	2573	2592	2601	2600	2590	2571	2543	2506	2461	5
5	1769	1863	1955	2044	2128	2207	2280	2347	2408	2461	4
6	0819	0900	0983	1070	1160	1253	1348	1445	1542	1641	3
7	0244	0279	0318	0360	0407	0458	0512	0571	0635	0703	2
8	0042	0051	0060	0071	0083	0097	0114	0132	0153	0176	1
9	0003	0004	0005	0006	0008	0009	0011	0014	0016	0020	0
	59	58	57	56	55	54	53	52	51	50 = p	x

n = 10

x	p = 01	02	03	04	05	06	07	08	09	10	
0	9044	8171	7374	6648	5987	5386	4840	4344	3894	3487	10
1	0914	1667	2281	2770	3151	3438	3643	3777	3851	3874	9
2	0042	0153	0317	0519	0746	0988	1234	1478	1714	1937	8
3	0001	0008	0026	0058	0105	0168	0248	0343	0452	0574	7
4	0000	0000	0001	0004	0010	0019	0033	0053	0078	0112	6
5	0000	0000	0000	0000	0001	0001	0003	0005	0009	0015	5
6	0000	0000	0000	0000	0000	0000	0000	0000	0001	0001	4
	99	98	97	96	95	94	93	92	91	90 = p	x

x	p = 11	12	13	14	15	16	17	18	19	20	
0	3118	2785	2484	2213	1969	1749	1552	1374	1216	1074	10
1	3854	3798	3712	3603	3474	3331	3178	3017	2852	2684	9
2	2143	2330	2496	2639	2759	2856	2929	2980	3010	3020	8
3	0706	0847	0995	1146	1298	1450	1600	1745	1883	2013	7
4	0153	0202	0260	0326	0401	0483	0573	0670	0773	0881	6
5	0023	0033	0047	0064	0085	0111	0141	0177	0218	0264	5
6	0002	0004	0006	0009	0012	0018	0024	0032	0043	0055	4
7	0000	0000	0000	0001	0001	0002	0003	0004	0006	0008	3
8	0000	0000	0000	0000	0000	0000	0000	0000	0001	0001	2
	89	88	87	86	85	84	83	82	81	80 = p	x

x	p = 21	22	23	24	25	26	27	28	29	30	
0	0947	0834	0733	0643	0563	0492	0430	0374	0326	0282	10
1	2517	2351	2188	2030	1877	1730	1590	1456	1330	1211	9
2	3011	2984	2942	2885	2816	2735	2646	2548	2444	2335	8
3	2134	2244	2343	2429	2503	2563	2609	2642	2662	2668	7
4	0993	1108	1225	1343	1460	1576	1689	1798	1903	2001	6
5	0317	0375	0439	0509	0584	0664	0750	0839	0933	1029	5
6	0070	0088	0109	0134	0162	0195	0231	0272	0317	0368	4
7	0011	0014	0019	0024	0031	0039	0049	0060	0074	0090	3
8	0001	0002	0002	0003	0004	0005	0007	0009	0011	0014	2
9	0000	0000	0000	0000	0000	0000	0001	0001	0001	0001	1
	79	78	77	76	75	74	73	72	71	70 = p	x

x	p = 31	32	33	34	35	36	37	38	39	40	
0	0245	0211	0182	0157	0135	0115	0098	0084	0071	0060	10
1	1099	0995	0898	0808	0725	0649	0578	0514	0456	0403	9
2	2222	2107	1990	1873	1757	1642	1529	1419	1312	1209	8
3	2662	2644	2614	2573	2522	2462	2394	2319	2237	2150	7
4	2093	2177	2253	2320	2377	2424	2461	2487	2503	2508	6
5	1128	1229	1332	1434	1536	1636	1734	1829	1920	2007	5
6	0422	0482	0547	0616	0689	0767	0849	0934	1023	1115	4
7	0108	0130	0154	0181	0212	0247	0285	0327	0374	0425	3
8	0018	0023	0028	0035	0043	0052	0063	0075	0090	0106	2
9	0002	0002	0003	0004	0005	0006	0008	0010	0013	0016	1
10	0000	0000	0000	0000	0000	0000	0000	0001	0001	0001	0
	69	68	67	66	65	64	63	62	61	60 = p	x

x	p = 41	42	43	44	45	46	47	48	49	50	
0	0051	0043	0036	0030	0025	0021	0017	0014	0012	0010	10
1	0355	0312	0273	0238	0207	0180	0155	0133	0114	0098	9
2	1111	1017	0927	0843	0763	0688	0619	0554	0494	0439	8
3	2058	1963	1865	1765	1665	1564	1464	1364	1267	1172	7
4	2503	2488	2462	2427	2384	2331	2271	2204	2130	2051	6
5	2087	2162	2229	2289	2340	2383	2417	2441	2456	2461	5
6	1209	1304	1401	1499	1596	1692	1786	1878	1966	2051	4
7	0480	0540	0604	0673	0746	0824	0905	0991	1080	1172	3
8	0125	0147	0171	0198	0229	0263	0301	0343	0389	0439	2
9	0019	0024	0029	0035	0042	0050	0059	0070	0083	0098	1
10	0001	0002	0003	0003	0003	0004	0005	0006	0008	0010	0
	59	58	57	56	55	54	53	52	51	50 = p	x

(Continued)

تابع ملحق (10)

Binomial distribution $(n = 20)$

$n = 20$

x	p	01	02	03	04	05	06	07	08	09	10		
0		8179	6676	5438	4420	3585	2901	2342	1887	1516	1216		20
1		1652	2725	3364	3683	3774	3703	3526	3282	3000	2702		19
2		0159	0528	0988	1458	1887	2246	2521	2711	2828	2852		18
3		0010	0065	0183	0364	0596	0860	1139	1414	1672	1901		17
4		0000	0006	0024	0065	0133	0233	0364	0523	0703	0898		16
5		0000	0000	0002	0009	0022	0048	0088	0145	0222	0319		15
6		0000	0000	0000	0001	0003	0008	0017	0032	0055	0089		14
7		0000	0000	0000	0000	0000	0001	0002	0005	0011	0020		13
8		0000	0000	0000	0000	0000	0000	0000	0001	0002	0004		12
9		0000	0000	0000	0000	0000	0000	0000	0000	0000	0001		11
		99	98	97	96	95	94	93	92	91	90	p	x

x	p	11	12	13	14	15	16	17	18	19	20		
0		0972	0776	0617	0490	0388	0306	0241	0189	0148	0115		20
1		2403	2115	1844	1595	1368	1165	0986	0829	0693	0576		19
2		2822	2740	2618	2466	2293	2109	1919	1730	1545	1369		18
3		2093	2242	2347	2409	2428	2410	2358	2278	2175	2054		17
4		1099	1299	1491	1666	1821	1951	2053	2125	2168	2182		16
5		0435	0567	0713	0868	1028	1189	1345	1493	1627	1746		15
6		0134	0193	0266	0353	0454	0566	0689	0819	0954	1091		14
7		0033	0053	0080	0115	0160	0216	0282	0360	0448	0545		13
8		0007	0012	0019	0030	0046	0067	0094	0128	0171	0222		12
9		0001	0002	0004	0007	0011	0017	0026	0038	0053	0074		11
10		0000	0000	0001	0001	0002	0004	0006	0009	0014	0020		10
11		0000	0000	0000	0000	0000	0001	0001	0002	0003	0005		9
12		0000	0000	0000	0000	0000	0000	0000	0000	0001	0001		8
		89	88	87	86	85	84	83	82	81	80	p	x

x	p	21	22	23	24	25	26	27	28	29	30		
0		0090	0069	0054	0041	0032	0024	0016	0014	0011	0008		20
1		0477	0392	0321	0261	0211	0170	0137	0109	0087	0068		19
2		1204	1050	0910	0783	0669	0569	0480	0403	0336	0278		18
3		1920	1777	1631	1484	1339	1199	1065	0940	0823	0716		17
4		2169	2131	2070	1991	1897	1790	1675	1553	1429	1304		16
5		1845	1923	1979	2012	2023	2013	1982	1933	1868	1789		15
6		1226	1356	1478	1589	1686	1768	1833	1879	1907	1916		14
7		0652	0765	0883	1003	1124	1242	1356	1462	1558	1643		13
8		0282	0351	0429	0515	0609	0709	0815	0924	1034	1144		12
9		0100	0132	0171	0217	0271	0332	0402	0479	0563	0654		11
10		0029	0041	0056	0075	0099	0128	0163	0203	0253	0308		10
11		0007	0010	0015	0022	0030	0041	0055	0072	0094	0120		9
12		0001	0002	0003	0005	0008	0011	0015	0021	0029	0039		8
13		0000	0000	0001	0001	0002	0002	0003	0005	0007	0010		7
14		0000	0000	0000	0000	0000	0000	0010	0010	0001	0002		6
		79	78	77	76	75	74	73	72	71	70	p	x

$n = 20$

x	p	31	32	33	34	35	36	37	38	39	40		
0		0006	0004	0003	0002	0002	0001	0001	0001	0001	0000		20
1		0054	0042	0033	0025	0020	0015	0011	0009	0007	0005		19
2		0229	0188	0153	0124	0100	0080	0064	0050	0040	0031		18
3		0619	0531	0453	0383	0323	0270	0224	0185	0152	0123		17
4		1181	1062	0947	0839	0738	0645	0559	0482	0412	0350		16
5		1698	1599	1493	1384	1272	1161	1051	0945	0843	0746		15
6		1907	1881	1839	1782	1712	1632	1543	1447	1347	1244		14
7		1714	1770	1811	1836	1844	1836	1812	1774	1722	1659		13
8		1251	1354	1450	1537	1614	1678	1730	1767	1790	1797		12
9		0750	0849	0952	1056	1158	1259	1354	1444	1526	1597		11
10		0370	0440	0516	0598	0686	0779	0875	0974	1073	1171		10
11		0151	0188	0231	0280	0336	0398	0467	0542	0624	0710		9
12		0051	0066	0085	0108	0136	0168	0206	0249	0299	0355		8
13		0014	0019	0026	0034	0043	0056	0074	0094	0118	0146		7
14		0003	0005	0006	0009	0012	0016	0022	0029	0038	0049		6
15		0001	0001	0001	0002	0003	0004	0005	0007	0010	0013		5
16		0000	0000	0000	0000	0000	0001	0001	0001	0002	0003		4
		69	68	67	66	65	64	63	62	61	60	p	x

x	p	41	42	43	44	45	46	47	48	49	50		
1		0004	0003	0002	0001	0001	0001	0001	0000	0000	0000		19
2		0024	0018	0014	0011	0008	0006	0005	0003	0002	0002		18
3		0100	0080	0064	0051	0040	0031	0024	0019	0014	0011		17
4		0295	0247	0206	0170	0139	0113	0092	0074	0059	0046		16
5		0656	0573	0496	0427	0365	0309	0260	0217	0180	0148		15
6		1140	1037	0936	0839	0746	0658	0577	0501	0432	0370		14
7		1585	1502	1413	1318	1221	1122	1023	0925	0830	0739		13
8		1790	1768	1732	1683	1623	1553	1474	1388	1296	1201		12
9		1658	1707	1742	1763	1771	1763	1742	1708	1661	1602		11
10		1268	1359	1446	1524	1593	1652	1700	1734	1755	1762		10
11		0801	0895	0991	1089	1185	1280	1370	1455	1533	1602		9
12		0417	0486	0561	0642	0727	0818	0911	1007	1105	1201		8
13		0178	0217	0260	0310	0366	0429	0497	0572	0653	0739		7
14		0062	0078	0098	0122	0150	0183	0221	0264	0314	0370		6
15		0017	0023	0030	0038	0049	0062	0078	0098	0121	0148		5
16		0004	0005	0007	0009	0013	0017	0022	0028	0036	0046		4
17		0001	0001	0001	0002	0002	0003	0005	0006	0008	0011		3
18		0000	0000	0000	0000	0000	0000	0001	0001	0001	0002		2
		59	58	57	56	55	54	53	52	51	50	p	x

(Continued)

ملحق (11) الأرقام العشوائية

31 75 15 72 60	68 98 00 53 39	15 47 04 83 55	88 65 12 25 96	03 15 21 91 21
88 49 29 93 82	14 45 40 45 04	20 09 49 89 77	74 84 39 34 13	22 10 97 85 08
30 93 44 77 44	07 48 18 38 28	73 78 80 65 33	28 59 72 04 05	94 20 52 03 80
22 88 84 88 93	27 49 99 87 48	60 53 04 51 28	74 02 28 46 17	82 03 71 02 68
78 21 21 69 93	35 90 29 13 86	44 37 21 54 86	65 74 11 40 14	87 48 13 72 20
41 84 98 45 47	46 85 05 23 26	34 67 75 83 00	74 91 06 43 45	19 32 58 15 49
46 35 23 30 49	69 24 89 34 60	45 30 50 75 21	61 31 83 18 55	14 41 37 09 51
11 08 79 62 94	14 01 33 17 92	59 74 76 72 77	76 50 33 45 13	39 66 37 75 44
52 70 10 83 37	56 30 38 73 15	16 52 06 96 76	11 63 49 98 93	02 18 16 81 61
57 27 53 68 98	81 30 44 85 85	68 65 22 73 76	92 85 25 58 66	88 44 80 35 84
20 85 77 31 56	70 28 42 43 26	79 37 59 52 20	01 15 96 32 67	10 62 24 83 91
15 63 38 49 24	90 41 59 36 14	33 52 12 66 65	55 82 34 76 41	86 22 53 17 04
92 69 44 82 97	39 90 40 21 15	59 58 94 90 67	66 82 14 15 75	49 76 70 40 37
77 61 31 90 19	88 15 20 00 80	20 55 49 14 09	96 27 74 82 57	50 81 69 76 16
38 68 83 24 86	45 13 46 35 45	59 40 47 20 59	43 94 75 16 80	43 85 25 96 93
25 16 30 18 89	70 01 41 50 21	41 29 06 73 12	71 85 71 59 57	68 97 11 14 03
65 25 10 76 29	37 23 93 32 95	05 87 00 11 19	92 78 42 63 40	18 47 76 56 22
36 81 54 36 25	18 63 73 75 09	82 44 49 90 05	04 92 17 37 01	14 70 79 39 97
64 39 71 16 92	05 32 78 21 62	20 24 78 17 59	45 19 72 53 32	83 74 52 25 67
04 51 52 56 24	95 09 66 79 46	48 46 08 55 58	15 19 11 87 82	16 93 03 33 61
15 88 09 22 61	17 29 28 81 90	61 78 14 88 98	92 52 52 12 83	88 58 16 00 98
71 92 60 08 19	59 14 40 02 24	30 57 09 01 94	18 32 90 69 99	26 85 71 92 38
64 42 52 81 08	16 55 41 60 16	00 04 28 32 29	10 33 33 61 68	65 61 79 48 34
79 78 22 39 24	49 44 03 04 32	81 07 73 15 43	95 21 66 48 65	13 65 85 10 81
36 33 77 45 38	44 55 36 46 72	90 96 04 18 49	93 86 54 46 08	93 17 63 48 51
05 24 92 93 29	19 71 59 40 82	14 73 88 66 67	43 70 86 63 54	93 69 22 55 27
56 46 39 93 80	38 79 38 57 74	19 05 61 39 39	46 06 22 76 47	66 14 66 32 10
96 29 63 31 21	54 19 63 41 08	75 81 48 59 86	71 17 11 51 02	28 99 26 31 65
98 38 03 62 69	60 01 40 72 01	62 44 84 63 85	42 17 58 83 50	46 18 24 91 26
52 56 76 43 50	16 31 55 39 69	80 39 58 11 14	54 35 86 45 78	47 26 91 57 47
78 49 89 08 30	25 95 59 92 36	43 28 69 10 64	99 96 99 51 44	64 42 47 73 77
49 55 32 42 41	08 15 08 95 35	08 70 39 10 41	77 32 38 10 79	45 12 79 36 86
32 15 10 70 75	83 15 51 02 52	73 10 08 86 18	23 89 18 74 18	45 41 72 02 68
11 31 45 03 63	26 86 02 77 99	49 41 68 35 34	19 18 70 80 59	76 67 70 21 10
12 36 47 12 10	87 05 25 02 41	90 78 59 78 89	81 39 95 81 30	64 43 90 56 14
09 18 82 00 97	32 82 53 95 27	04 22 08 63 04	83 38 98 73 74	64 27 85 80 44
90 04 58 54 97	51 98 15 06 54	94 93 88 19 97	91 87 07 61 50	68 47 66 46 59
73 18 95 02 07	47 67 72 62 69	62 29 06 44 64	27 12 46 70 18	41 36 18 27 60
75 76 87 64 90	20 97 18 17 49	90 42 91 22 72	95 37 50 58 71	93 82 34 31 78
54 01 64 40 56	66 28 13 10 03	00 68 22 73 98	20 71 45 32 95	07 70 61 78 13
08 35 86 99 10	78 54 24 27 85	13 66 15 88 73	04 61 89 75 53	31 22 30 84 20
28 30 60 32 64	81 33 31 05 91	40 51 00 78 93	32 60 46 04 75	94 11 90 18 40
53 84 08 62 33	81 59 41 36 28	51 21 59 02 90	28 46 66 87 95	77 76 22 07 91
91 75 75 37 41	61 61 36 22 69	50 26 39 02 12	55 78 17 65 14	83 48 34 70 55
89 41 59 26 94	00 39 75 83 91	12 60 71 76 46	48 94 97 23 06	94 54 13 74 08
77 31 30 38 20	88 83 42 99 01	68 41 48 27 74	51 90 81 39 80	72 89 35 55 07
19 50 23 71 74	69 97 92 02 88	55 21 02 97 73	74 28 77 52 51	65 34 46 74 15
21 81 85 93 13	93 27 88 17 57	05 68 67 31 56	07 08 28 50 46	31 85 33 84 52
51 47 46 64 99	68 10 72 36 21	94 04 99 13 45	42 83 60 91 91	08 00 74 54 49
99 55 96 83 31	62 53 52 41 70	69 77 71 28 30	74 81 97 81 42	43 86 07 28 34
60 31 14 28 24	37 30 14 26 78	45 99 04 32 42	17 37 45 20 03	70 70 77 02 14
49 73 97 14 84	92 00 39 80 86	76 66 87 32 09	59 20 21 19 73	02 90 23 32 50
78 62 65 15 94	16 45 39 46 14	39 01 49 70 66	83 01 20 98 32	25 57 17 76 28
66 69 21 39 86	99 83 70 05 82	81 23 24 49 87	09 50 49 64 12	90 19 37 95 68
44 07 12 80 91	07 36 29 77 03	76 44 74 25 37	98 52 49 78 31	65 70 40 95 14
41 46 88 51 49	49 55 41 79 94	14 92 43 96 50	95 29 40 05 56	70 48 10 69 05
94 55 93 75 59	49 67 85 31 19	70 31 20 56 82	66 98 63 40 99	74 47 42 07 40
41 61 57 03 60	64 11 45 86 60	90 85 06 46 18	80 62 05 17 90	11 43 63 80 72
50 27 39 31 13	41 79 48 68 61	24 78 18 96 83	55 41 18 56 67	77 53 59 98 92
41 39 68 05 04	90 67 00 82 89	40 90 20 50 69	95 08 30 67 83	28 10 25 78 16

تابع ملحق (11)

25	80	72	42	60	71	52	97	89	20	72	68	20	73	85	90	72	65	71	66	98	88	40	85	83
06	17	09	79	65	88	30	29	80	41	21	44	34	18	08	68	98	48	36	20	89	74	79	88	82
60	80	85	44	44	74	41	28	11	05	01	17	62	88	38	36	42	11	64	89	18	05	95	10	61
80	94	04	48	93	10	40	83	62	22	80	58	27	19	44	92	63	84	03	33	67	05	41	60	67
19	51	69	01	20	46	75	97	16	43	13	17	75	52	92	21	03	68	28	08	77	50	19	74	27
49	38	65	44	80	23	60	42	35	54	21	78	54	11	01	91	17	81	01	74	29	42	09	04	38
06	31	28	89	40	15	99	56	93	21	47	45	86	48	09	98	18	98	18	51	29	65	18	42	15
60	94	20	03	07	11	89	79	26	74	40	40	56	80	32	96	71	75	42	44	10	70	14	13	93
92	32	99	89	32	78	28	44	63	47	71	20	99	20	61	39	44	89	31	36	25	72	20	85	64
77	93	66	35	74	31	38	45	19	24	85	66	12	96	71	58	13	71	78	20	22	75	13	65	18
91	30	70	69	91	19	07	22	42	10	36	69	95	37	28	28	82	53	57	93	28	97	66	62	52
68	43	49	46	88	84	47	31	36	22	62	12	69	84	08	12	84	38	25	90	09	81	59	31	46
48	90	81	58	77	54	74	52	45	91	35	70	00	47	54	83	82	45	26	92	54	13	05	51	60
06	91	34	51	97	42	67	27	86	01	11	88	30	95	28	63	01	19	89	01	14	97	44	03	44
10	45	51	60	19	14	21	03	37	12	91	34	23	78	21	88	32	58	08	51	43	66	77	08	83
12	88	39	73	43	65	02	76	11	84	04	28	50	13	92	17	97	41	50	77	90	71	22	67	69
21	77	83	09	76	38	80	73	69	61	31	64	94	20	96	63	28	10	20	23	08	81	64	74	49
19	52	35	95	15	65	12	25	96	59	86	28	36	82	58	69	57	21	37	98	16	43	59	15	29
67	24	55	26	70	35	58	31	65	63	79	24	68	66	86	76	46	33	42	22	26	65	59	08	02
60	58	44	73	77	07	50	03	79	92	45	13	42	65	29	26	76	08	36	37	41	32	64	43	44
53	85	34	13	77	36	06	69	48	50	58	83	87	38	59	49	36	47	33	31	96	24	04	36	42
24	63	73	87	36	74	38	48	93	42	52	62	30	79	92	12	36	91	86	01	03	74	28	38	73
83	08	01	24	51	38	99	22	28	15	07	75	95	17	77	97	37	72	75	85	51	97	23	78	67
16	44	42	43	34	36	15	19	90	73	27	49	37	09	39	85	13	03	25	52	54	84	65	47	59
60	79	01	81	57	57	17	86	57	62	11	16	17	85	76	45	81	95	29	79	65	13	00	48	60
94	01	54	68	74	32	44	44	82	77	59	82	09	61	63	64	65	42	58	43	41	14	54	28	20
74	10	88	82	22	88	57	07	40	15	25	70	49	10	35	01	75	51	47	50	48	96	83	86	03
62	88	08	78	73	95	16	05	92	21	22	30	49	03	14	72	87	71	73	34	39	28	30	41	49
11	74	81	21	02	80	58	04	18	67	17	71	05	96	21	06	55	40	78	50	73	95	07	95	52
17	94	40	56	00	60	47	80	33	43	25	85	25	89	05	57	21	63	96	18	49	85	69	93	26
66	06	74	27	92	95	04	35	26	80	46	78	05	64	87	09	97	15	94	81	37	00	62	21	86
54	24	49	10	30	45	54	77	08	18	59	84	99	61	69	61	45	92	16	47	87	41	71	71	98
30	94	55	75	89	31	73	25	72	60	47	67	00	76	54	46	37	62	53	66	94	74	64	95	80
69	17	03	74	03	86	99	59	03	07	94	30	47	18	03	26	82	50	55	11	12	45	99	13	14
08	34	58	89	75	35	84	18	57	71	08	10	55	99	87	87	11	22	14	76	14	71	37	11	81
27	76	74	35	84	85	30	18	89	77	29	49	06	97	14	73	03	54	12	07	74	69	90	93	10
13	02	51	43	38	54	06	61	52	43	47	72	46	67	33	47	43	14	39	05	31	04	85	66	99
80	21	73	62	92	98	52	52	43	35	24	43	22	48	96	43	27	75	88	74	11	46	61	60	82
10	87	56	20	04	90	39	16	11	05	57	41	10	63	68	53	85	63	07	43	08	67	08	47	41
54	12	75	73	26	26	62	91	90	87	24	47	28	87	79	30	54	02	78	86	61	73	27	54	54
33	71	34	80	07	93	58	47	28	69	51	92	66	47	21	58	30	32	98	22	93	17	49	39	72
85	27	48	68	93	11	30	32	92	70	28	83	43	41	37	73	51	59	04	00	71	14	84	36	43
84	13	38	96	40	44	03	55	21	66	73	85	27	00	91	61	22	26	05	61	62	32	71	84	23
56	73	21	62	34	17	39	59	61	31	10	12	39	16	22	85	49	65	75	60	81	60	41	88	80
65	13	85	68	06	87	64	88	52	61	34	31	36	58	61	45	87	52	10	69	85	64	44	72	77
38	00	10	21	76	81	71	91	17	11	71	60	29	29	37	74	21	96	40	49	65	58	44	96	98
37	40	29	63	97	01	30	47	75	86	56	27	11	00	86	47	32	46	26	05	40	03	03	74	38
97	12	54	03	48	87	08	33	14	17	21	81	53	92	50	75	23	76	20	47	15	50	12	95	78
21	82	64	11	34	47	14	33	40	72	64	63	88	59	02	49	13	90	64	41	03	85	65	45	52
73	13	54	27	42	95	71	90	90	35	85	79	47	42	96	08	78	98	81	56	64	69	11	92	02
07	63	87	79	29	03	06	11	80	72	96	20	74	41	56	23	82	19	95	38	04	71	36	69	94
60	52	88	34	41	07	95	41	98	14	59	17	52	06	95	05	53	35	21	39	61	21	20	64	55
83	59	63	56	55	06	95	89	29	83	05	12	80	97	19	77	43	35	37	83	92	30	15	04	98
10	85	06	27	46	99	59	91	05	07	13	49	90	63	19	53	07	57	18	39	06	41	01	93	62
39	82	09	89	52	43	62	26	31	47	64	42	18	08	14	43	80	00	93	51	31	02	47	31	67
59	58	00	64	78	75	56	97	88	00	88	83	55	44	86	23	76	80	61	56	04	11	10	84	08
38	50	80	73	41	23	79	34	87	63	90	82	29	70	22	17	71	90	42	07	95	95	44	99	53
30	69	27	06	68	94	68	81	61	27	56	19	68	00	91	82	06	76	34	00	05	46	26	92	00
65	44	39	56	59	18	28	82	74	37	49	63	22	40	41	08	33	76	56	76	96	29	99	08	36
27	26	75	02	64	13	19	27	22	94	07	47	74	46	06	17	98	54	89	11	97	34	13	03	58

تابع ملحق (11)

38	10	17	77	56	11	65	71	38	97	95	88	95	70	67	47	64	81	38	85	70	66	99	34	06
39	64	16	94	57	91	33	92	25	02	92	61	38	97	19	11	94	75	62	03	19	32	42	05	04
84	05	44	04	55	99	39	66	36	80	67	66	76	06	31	69	18	19	68	45	38	52	51	16	00
47	46	80	35	77	57	64	96	32	66	24	70	07	15	94	14	00	42	31	53	69	24	90	57	47
43	32	13	13	70	28	97	72	38	96	76	47	96	85	62	62	34	20	75	89	08	89	90	59	85
64	28	16	18	26	18	55	56	49	37	13	17	33	33	65	78	85	11	64	99	87	06	41	30	75
66	84	77	04	95	32	35	00	29	85	86	71	63	87	46	26	31	37	74	63	55	38	77	26	81
72	46	13	32	30	21	52	95	34	24	92	58	10	22	62	78	43	86	62	76	18	39	67	35	38
21	03	29	10	50	13	05	81	62	18	12	47	05	65	00	15	29	27	61	39	59	52	65	21	13
95	36	26	70	11	06	65	11	61	36	01	01	60	08	57	55	01	85	63	74	35	82	47	17	08
40	71	29	73	80	10	40	45	54	52	34	03	06	07	26	75	21	11	02	71	36	63	36	84	24
58	27	56	17	64	97	58	65	47	16	50	25	94	63	45	87	19	54	60	92	26	78	76	09	39
89	51	41	17	88	68	22	42	34	17	73	95	97	61	45	30	34	24	02	77	11	04	97	20	49
15	47	25	06	69	48	13	93	67	32	46	87	43	70	88	73	46	50	98	19	58	86	93	52	20
12	12	08	61	24	51	24	74	43	02	60	88	35	21	09	21	43	73	67	86	49	22	67	78	37
03	99	11	04	61	93	71	61	68	94	66	08	32	46	53	84	60	95	82	32	88	61	81	91	61
38	55	59	55	54	32	88	65	97	80	08	35	56	08	60	29	73	54	77	62	71	29	92	38	53
17	54	67	37	04	92	05	24	62	15	55	12	12	92	81	59	07	60	79	36	27	95	45	89	09
32	64	35	28	61	95	81	90	68	31	00	91	19	89	36	76	35	59	37	79	80	86	30	05	14
69	57	26	87	77	39	51	03	59	05	14	06	04	06	19	29	54	96	96	16	33	56	46	07	80
24	12	26	65	91	27	69	90	64	94	14	84	54	66	72	61	95	87	71	00	90	89	97	57	54
61	19	63	02	31	92	96	26	17	73	41	83	95	53	82	17	26	77	09	43	78	03	87	02	67
30	53	22	17	04	10	27	41	22	02	39	68	52	33	09	10	06	16	88	29	55	98	66	64	89
03	78	89	75	99	75	86	72	07	17	74	41	65	31	66	35	20	83	33	74	87	53	90	88	23
48	22	86	33	79	85	78	34	76	19	53	15	26	74	33	35	66	35	29	72	16	81	86	03	11
60	36	59	46	53	35	07	53	39	49	42	61	42	92	97	01	91	82	83	16	98	95	37	32	31
83	79	94	24	02	56	62	33	44	42	34	99	44	13	74	70	07	11	47	36	09	95	81	80	65
32	96	00	74	05	36	40	98	32	32	99	38	54	16	00	11	13	30	75	86	15	91	70	62	53
19	32	25	38	45	57	62	05	26	06	66	49	76	86	46	78	13	86	65	59	19	64	09	94	13
11	22	09	47	47	07	39	93	74	08	48	50	92	39	29	27	48	24	54	76	85	24	43	51	59
21	44	58	27	93	24	83	19	32	41	14	19	97	62	68	70	88	36	80	02	03	82	91	74	43
72	51	37	64	00	52	22	59	23	48	62	30	89	84	81	29	74	43	31	65	33	14	16	10	20
71	47	94	50	27	76	16	05	74	11	13	78	01	36	32	52	30	87	77	62	88	87	43	36	97
83	21	05	14	66	09	08	85	03	95	26	74	30	53	06	21	70	67	00	01	99	43	98	07	67
68	74	99	51	48	94	89	77	86	36	96	75	00	90	24	94	53	89	11	43	96	39	36	18	86
05	18	47	57	63	47	07	58	81	58	05	31	35	34	39	14	90	80	88	30	60	09	62	15	51
13	65	16	25	46	96	89	22	52	40	47	51	15	84	83	87	34	27	88	18	07	85	53	92	69
00	56	62	12	20	00	29	22	40	69	25	07	22	95	19	52	54	85	40	91	21	28	22	12	96
50	95	81	76	95	58	07	26	89	90	60	32	99	59	55	71	58	66	34	17	35	94	76	78	07
57	62	16	45	47	46	85	03	79	81	38	52	70	90	37	64	75	60	33	24	04	98	68	36	66
09	28	22	58	44	79	13	97	84	35	35	42	84	35	61	69	79	96	33	14	12	99	19	35	16
23	39	49	42	06	93	43	23	78	36	94	91	92	68	46	02	55	57	44	10	94	91	54	81	99
05	28	03	74	70	93	62	20	43	45	15	09	21	95	10	18	09	41	66	13	78	23	45	00	01
95	49	19	79	76	38	30	63	21	92	82	63	95	46	24	72	43	49	26	06	23	19	17	46	93
78	52	10	01	04	18	24	87	55	83	90	32	65	07	85	54	03	46	62	51	35	77	41	46	92
96	34	54	45	79	86	93	24	40	53	75	70	42	08	40	86	58	38	39	44	52	45	67	37	66
77	96	33	11	51	32	36	49	16	91	47	35	74	03	38	23	43	52	40	65	08	45	89	53	66
07	52	01	12	94	23	23	80	17	48	41	69	06	73	28	54	81	43	77	77	10	05	74	23	32
38	42	30	23	09	70	70	38	57	36	46	14	81	42	58	29	23	61	21	52	05	08	86	58	25
02	46	36	55	33	21	19	96	05	55	33	92	80	18	17	07	39	68	92	15	30	72	22	21	02
83	76	16	08	73	43	25	38	41	45	60	83	32	59	83	01	29	14	13	49	20	36	80	71	26
14	38	70	63	45	80	85	40	92	79	43	52	90	63	18	38	38	47	47	61	41	19	63	74	80
51	32	19	22	46	80	08	87	70	74	88	72	25	67	36	66	16	44	94	31	66	91	93	16	78
72	47	20	00	08	80	89	01	80	02	94	81	33	19	00	54	15	58	34	36	35	35	25	41	31
05	46	65	53	06	93	12	81	84	64	74	45	79	05	61	72	84	81	18	34	79	98	26	84	16
39	52	87	24	84	82	47	42	55	93	48	54	53	52	47	18	61	91	36	74	18	61	11	92	41
81	61	61	87	11	53	34	24	42	76	75	12	21	17	24	74	62	77	37	07	58	31	91	59	97
07	58	61	61	20	82	64	12	28	20	88	90	41	31	41	32	39	21	97	63	61	19	96	79	40
90	76	70	42	35	13	57	41	72	00	69	90	26	37	42	78	46	42	25	01	18	62	79	08	72
40	18	82	81	93	29	59	38	86	27	94	97	21	15	98	62	09	53	67	87	00	44	15	89	97

تابع ملحق (11)

34	41	48	21	57	86	88	75	50	87	19	15	20	00	23	12	30	28	07	83	32	62	46	86	91
63	43	97	53	63	44	98	91	68	22	36	02	40	08	67	76	37	84	16	05	65	96	17	34	88
67	04	90	90	70	93	39	94	55	47	94	45	87	42	84	05	04	14	98	07	20	28	83	40	60
79	49	50	41	46	52	16	29	02	86	54	15	83	42	43	46	97	83	54	82	59	36	29	59	38
91	70	43	05	52	04	73	72	10	31	75	05	19	30	29	47	66	56	43	82	99	78	29	34	78
19	61	27	84	30	11	66	19	47	70	77	60	36	56	69	86	86	81	26	65	30	01	27	59	89
39	14	17	74	00	28	00	06	42	38	73	25	87	17	94	31	34	02	62	56	66	45	33	70	16
64	75	68	04	57	08	74	71	28	36	03	46	95	06	78	03	27	44	34	23	66	67	78	25	56
92	90	15	18	78	56	44	12	29	98	29	71	83	84	47	06	45	32	53	11	07	56	55	37	71
03	55	19	00	70	09	48	39	40	50	45	93	81	81	35	36	90	84	33	21	11	07	35	18	03
98	88	46	62	09	06	83	05	36	56	14	66	35	63	46	71	43	00	49	09	19	81	80	57	07
27	36	98	68	82	53	47	30	76	41	53	63	37	08	63	03	74	81	28	22	19	36	04	90	88
59	06	67	59	74	63	33	52	04	83	43	51	43	74	81	58	27	82	69	67	49	43	54	39	51
91	64	79	37	83	64	16	94	90	22	98	58	80	94	95	49	82	95	90	68	38	83	10	48	38
83	60	59	24	19	39	54	20	77	72	71	56	87	56	73	35	18	58	97	59	44	90	17	42	91
24	89	58	85	30	70	77	43	54	39	46	75	87	04	72	70	20	79	26	75	91	62	36	12	75
35	72	02	65	56	95	59	62	00	94	73	75	08	57	88	34	26	40	17	03	46	83	36	52	48
14	14	15	34	10	38	64	90	63	43	57	25	66	13	42	72	70	97	53	18	90	37	93	75	62
27	41	67	56	70	92	17	67	25	35	93	11	95	60	77	06	88	61	82	44	92	34	43	13	74
82	07	10	74	29	81	00	[illegible]	77	49	40	74	45	69	74	23	33	68	88	21	53	84	11	05	36

المراجع العلمية

1. Introduction to Statistics, Third Ed., Ronald E. Waqlpole
(collier MacMillan, 1982: الناشر)

2. Introduction to Statistical Quality Control, Douglas C. Moutgouery
(JohnWiley, 1985: الناشر)

3. Experimental designs, Second Ed., W. G Cochran and G.M.Cox
(A Wiley, 1957: الناشر)

4. Elementary Statistics, Fourth Ed., Paul G. Hoel
(A Wiley Arabook, 1984: الناشر)

5. Experimental Design, Theory and Application, Walter. T. Fcderer
(Oxford and IBH, 1955: الناشر)

6. Elements of Statistical inference 4^{th} ed., by David V. Huntsberger and Patrick Billingsley
(Allyn and Bacon, 1977: الناشر)

7. Statistics and Experimental Design in Engineering and physical Sciences- 2ed. Ed, by Norman L.Johnson and Fred C. Leone
(John Willey, 1977: الناشر)

8. Mathematical Statistics, by Md. Fasihnddin
(Library Promotion Bureau Karachi-2-, 1968: الناشر)

9. Introductory Statistical Analysis, 2ed ed., by Donald L. Harnett and James L. Murphy
(Addison Wesley, 1980: الناشر)

10. Mathematical Statistics with Applications by W. Mendenhall and R. Scheaffer
(DuxBury Press, 1973: الناشر)

11. Introduction to Probability and Statistics, Principles and Applications for Engineering and the Computing Sciences Second Ed., by J. S. Milton and Jesse C. Arnold

(McGraw-Hill Publishing Company, 1990: الناشر)

12. Matematiksel Istatistige Giris, U. Korum

(Ankara.U, 1985: الناشر)

13. Probability, by Symour Lipschntz Schum's Series

(MacGraw hill co., 1974: الناشر)

14- أوليات الإحصاء، الدكتور القبانجي، كلية الاقتصاد – جامعة قاريونس 1985

15- الاحتمالات، الدكتور صلاح الأحمد، كلية العلوم – جامعة دمشق – 1982.

16- الاحتمالات، الدكتور عدنان عمورة، كلية العلوم – جامعة دمشق – 1993.

17- مبادئ الإحصاء والاحتمالات، الدكتور أنور اللحام والدكتور محمد شفيق ياسين – كلية العلوم – جامعة دمشق 1982.

18- الاحتمال والإحصاء الحيوي، الدكتور علي سلامي والدكتور سلطان الصلخدي، كلية الهندسة

اللجنة العلميّة

أ. د. علاء الدين القبانجي.

أ. م. د. عدنان عمورة.

أ. م. د. حسان العاقل.

المدقق اللغوي

د. مزيد نعيم.

حقوق الطبع والترجمة والنشر محفوظة لمديرية الكتب والمطبوعات

بسم الله الرحمن الرحيم

كتاب الاحتمال والإحصاء

نسخة إلكترونية مهداة من

مكتبة السماح للخدمات الجامعية ((أبو وائل))

لطلب نسخة منسوخة في حال عدم توفر النسخة الأصلية

يرجى الاتصال على 0933467204

أو تشريفنا على عنواننا

قريب من السكن الجامعي – مقابل حديقة أبناء الشهداء

أول جادة المعتصم 10م

لا تنسونا من صالح الدعاء

والله ولي التوفيق